GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 21:10:52 ; Search time 212.959 Seconds
(without alignments)
16512.552 Million cell updates/sec

Title: US-10-027-000-1
Perfect score: 2976
Sequence: 1 ttatagtcgctttgttaat.........aaaaaaaaaaaaaaaaa 2976

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues
Total number of hits satisfying chosen parameters: 1492128

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB	ID	Description
1	175.2	5.9	2430	9	US-09-860-846-23	Sequence 23, Appl
2	175.2	5.9	2430	10	US-09-988-384B-23	Sequence 23, Appl
3	175.2	5.9	2430	10	US-09-861-289-23	Sequence 23, Appl
4	175.2	5.9	12441	9	US-09-988-384B-3	Sequence 3, Appli
5	175.2	5.9	13613	9	US-09-861-289-3	Sequence 3, Appli
6	175.2	5.9	13613	10	US-09-861-289-3	Sequence 3, Appli
7	138.6	4.7	50937	9	US-09-808-880-1	Sequence 1, Appli
8	138.6	4.6	2166	9	US-09-121-032-5	Sequence 5, Appl
c 9	135.8	4.6	2166	9	US-10-093-037-5	Sequence 5, Appli
10	135.8	4.6	390	10	US-09-790-399-7	Sequence 7, Appl
11	84.4	2.8	390	10	US-09-748-033-2	Sequence 2, Appl
12	78.2	2.6	1294	10	US-09-748-033-4	Sequence 4, Appl
13	76.2	2.6	1107	10	US-09-748-033-6	Sequence 6, Appli
14	76.2	2.6	2712	10	US-09-815-242-7960	Sequence 7960, Ap
15	73.6	2.5	1914	10	US-09-476-242-9	Sequence 9, Appli
16	73.2	2.5	2541	10	US-09-887-576-784	Sequence 784, App
17	73	2.5	1185	12	US-10-032-717-9	Sequence 9, Appl
18	71.6	2.4	2010	12	US-09-476-242-10	Sequence 10, Appl
19	71.6	2.4	2541	10	US-09-476-242-12	Sequence 12, Appl
20	70.8	2.4	1929	9	US-09-899-642-1	Sequence 1, Appl
21	70.8	2.4	13842	9	US-09-860-846-30	Sequence 30, Appl
22	70.8	2.4	13842	9	US-09-988-384B-30	Sequence 30, Appl
23	70.8	2.4	13842	10	US-09-861-289-30	Sequence 30, Appl
24	70.8	2.4	36778	10	US-09-860-846-5	Sequence 5, Appli
25	70.8	2.4	36778	10	US-09-861-289-5	Sequence 5, Appli
26	70.8	2.4	37948	10	US-09-988-384B-5	Sequence 5, Appli
27	70	2.4	2541	10	US-09-476-242-11	Sequence 11, Appl
28	69.8	2.3	88421	9	US-09-976-059-1	Sequence 1, Appl
c 29	68.8	2.3	1896	9	US-10-124-880-162	Sequence 162, App
30	68	2.3	804	9	US-09-864-864-162	Sequence 15, App
31	67.6	2.3	1512	9	US-09-773-748-2	Sequence 2, Appl
32	67.6	2.3	15872	9	US-10-124-800-5	Sequence 5, Appl
c 33	67.6	2.3	15872	10	US-09-988-384B-1	Sequence 1, Appli
34	67.4	2.3	310	10	US-09-861-289-1	Sequence 1, Appli
35	67	2.3	1896	10	US-10-124-800-27	Sequence 27, Appl
36	67	2.3	4512	9	US-09-960-352-3525	Sequence 3525, Ap
37	66.8	2.3	4509	9	US-09-748-033-7	Sequence 7, Appli
38	66.8	2.2	434	10	US-09-988-384B-1	Sequence 31, Appl
39	66.6	2.2	905	10	US-09-826-660-26	Sequence 26, Appl
40	66.6	2.2	1965	10	US-09-476-242-20	Sequence 20, Appl
41	66.6	2.2	2322	10	US-09-988-462-2	Sequence 2, Appli
42	66.4	2.2	3468	9	US-09-878-574-4765	Sequence 4765, Ap
43	66.2	2.2	495	10	US-09-918-995-5818	Sequence 5818, Ap
44	66	2.2	429	10	US-09-960-352-5390	Sequence 5390, Ap
45	65.8	2.2	362	10	US-09-960-352-5390	Sequence 5390, Ap

ALIGNMENTS

RESULT 1
US-09-860-846-23 Sequence 23, Application US/09860846
Sequence 23, Application US/09860846
Patent No. US20020164742A1

GENERAL INFORMATION:
APPLICANT: Sherman, D. H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/860,846
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 23
LENGTH: 2430
TYPE: DNA
ORGANISM: Streptomyces venezuelae

US-09-860-846-23

Query Match 47.6%; Score 175.2; DB 9; Length 2430;
Best Local Similarity 47.6%;
Matches 1016; Conservative 0; Mismatches 968; Indels 150; Gaps 11;

```
Qy  544 CGCTCGGTTCCACATTCAACCAAACTCTGTCGAAGAGGCCAGGTAAGATGAATGGGCAAAG 603
            |||| |  ||  ||  |  ||  || ||  ||  ||  |||  ||  | || | |||  |||
Db  353 CCCTGGCCAGCACCTTCGACGACCACCATGGCCGCGACAGCTPACGGCAAAGGTCATGGGCCG 412

Qy  604 AGGCCATCGCTAAGAGCTGCCGCATTGTCGATCCTTCGGCCCGACTATCAAACATGCAAGCGCTCCC 663
            | |  | ||  |  || ||| ||  ||| || ||| |||  ||| ||  ||  || ||  ||  |
Db  413 ACGGTCGCGCCGCCTCAACCAGGACATGGTCGTTCCTGGGCCCGATTGATGAACAACATCCGGGTGC 472

Qy  664 CTCTCGGTTGGACGTGGCGTTCGGAGTCGATTGGTGAGGATCCGTTCCTTGGGCGGGGCTTGGGAG 723
            | | |||  ||  ||| ||  | | || || ||| |||| ||  || ||| |  ||| |||  |
Db  473 CGCACGGCGGCGCGGAGACTACCGAGACCTTCACGCAGACCCCCTGGTCTCCGTCGGCGACTTCG 532

Qy  724 CTGCGCGGTCTCATCCCGCGGGGCATTCAGAGCACTGGAGTGCAGGCTACGGATCAAAGCACTTTT 783
            | | ||| ||| || || |||| || |||| ||| || || || |||||  | || | || || |
Db  533 CGGTCGCGCCAGATCAAGGGCATCACCAAGGGGTCGGTCTGATGACCACGGCCAAGCACTTCG 592
```

```
Qy    784 TGTCCAATGATCAGGAGGACAGCGCATGATGATGGTGCAGAGCAGCGGGCTC 843
              ||| |||  ||||  ||| ||||  ||| || ||  ||||| ||||||
Db    593 CGGCCAACAACAGGAGAACAACCGCTTTCTCCGTGAACGCCAAATGTCGACGCAGACGC 652

Qy    844 TCCGTGAAATCTACGACACTCCCGTTCCGAGATTGCTGTGTGACAGCACGGGTGCGT 903
              |||  ||||  || || ||||  ||||||  |||  ||||||   ||||||||
Db    653 TCCGCGACATCGAGTTCCCGGCGGTTCGAG---GCGTCCTCCAAGGCGGCGGGCTCCT 709

Qy    904 TCATGACGGCGGTACAATGGCATCAATGGCGTGCGAGAAACCTTAAAATATCTTG 963
              ||| |||||||  | || ||  |||||||  |||||||| ||| ||
Db    710 TCATGTGTCGTACTACAACGGCTCTCAACGGGATGGCGGAAGCCGTACGAGACGTCCTCA 769

Qy    964 ATGGGATGCTTCGAAAGGAATTAGGGGTTGGGACTGGGTACGGCA 1023
              |||  ||||||||  ||||| || ||||||||||||  ||||||||
Db    770 ACAACGGTGCTCGGCACGCAGTGGGGCTTCCAGGGCGGACCTGCTCGCCA 829

Qy   1024 CATACAGTACCACAGAAGCGGTGTGACCTCGAGATGCCGGACCTCCAC 1083
              |||  |||||||||   ||||||||||||||| ||||||
Db    830 C---CCCGGGCACCGAACCATCACCAAGGGCCTCGACCAGGAGATG--- 873

Qy   1084 GCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCTTTATCCACGTCA 1143
              ||||  ||||| || ||||||| || |||||  ||||||||||| || |||||||||
Db    874 ------GGCTGCAGCTCCCCGGCGTCCGGGACGTCCCGGAGGGCTGCACCAGGAGATG 925

Qy   1144 TTGACCAGAGGGGCTAGGGAAGTTCTTCAGTTCGTGCTGCTGCCCTCCGGAGTGA 1203
              |||||||  |||||||||||||||| ||||  || ||||  |||||||||||||
Db    926 CCGTGACGGGGTCGGGGTCGGGGATCGGAGAACGGCCACGGTCCCCAGGGCGGG 982

Qy   1204 CGGAGAACGGCCCCGAGACGACTGTCACACACACGGCAGCTCTCCTCCGGA 1263
              |||  ||||||||||  |||| ||||||| ||||||  |||||||||||
Db    983 CCGGTGAACGGGGTCGGGGGATCGTTGGGCGCCAAGGGTTCCCGGGGCGGG 1042

Qy   1264 AGGTTGGCAACGAGGGGCATCGTGCTGCTGTGAAGAACGTTCTGCCCTTGAGCA 1323
              |||||||||||||||||||| ||  |||||||||||||| ||||||||||||||
Db   1043 CCACTCGCGGGCGCGGGCGGCTGACTGGTGCGGACGCGGTGTCCCGAAGG 1102

Qy   1324 AGAAGAAGAAGAACGGCTGATTGTGCGGGCCCCCAAGCGCGGGTGCCCAAGG 1383
              |||   |||||||||||||| |||||||||||||||||||||||||  ||||
Db   1103 TCGCCGAGAAACGGCGGCGGTTCCTGCCCAAGCAGCCCTTCCGCGCCGGGTG 1162

Qy   1384 GCTTGCGCCACTCAGGGCGCCTACTACGGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGC 1443
              ||| |||||||||||||||||| |||||||||||||||  ||||||||||||
Db   1163 ACGCGGGAAGACCATCGCGGTCGCATCGGCCTCAGCCCCGTCGGAGGCAACGC 1221

Qy   1444 TCGAGAGACCGCGCCGCCATCGTACACCGTCGGGGCCGCCTTCCTCCCATTCTTAGGCG 1503
              |||||||||||||| |||||||||||||  ||||||||||||||||||| ||| ||||
Db   1222 CTGGGCACGCGCCACGTTGCTCCGGTCCCCGACTCGGACACCAACAAGGCC 1281

Qy   1504 AGCAGTGCCTCACGCCCGACGGCGCGCCATCGCGGGCGCTGGAGGGTGTCTTCAACGAGCCC 1563
              |||| ||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1282 CCGCGCGGTGGGGCGCCGGTGACGTACAGAACGGGTGCCTACGGCGGCAG 1341

Qy   1564 CTGGTACCCCTAACCGCCCAGACACATTGACGAGCTCTTCTTCACCAAGACGGACAATGCACC 1623
              ||| ||||||||||||||||  ||||||||||||  || ||||||
Db   1342 ATCCCGCGCGGGGAACCTCAGC----------------CCGGCGTTCACC 1375

Qy   1624 TCGTTGGACTACTACCACCCCAAAGGGGGCAGACACGTGTTACGGCCAGCACGT 1683
              |||| |||||||||||||  ||| |||||||| ||||||||||||||||||||
Db   1376 AGGGCCACCAGCTCGAGCCGGGCAAGGCGGGCCGCTGTACGACGCACGCTGACCGTGC 1435

Qy   1684 ACACCCGCGACGAGGACTGCACCTACGACGTCGGCCTCGTGTCGCTGCCGACGGCAAAG 1743
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1436 CCGCCGACGCGGAGTTACCGCCATCCGGTGTTACGCCCACACCACGGTGCAAAG 1494

Qy   1744 CGTACGTTAGACGACCAGCTCGTTCGTCGACAACGCCACCAGTCCCGGATGCGCT 1803
              ||| |||||||||||||||||||||||||
Db   1495 ------------------------------CTCGGCGAGGCCACA 1507

Qy   1804 TCTTCGGCTCCGCCACCCCGCGAGGAGACGGGGCCGATGATGCAGAGCACGT 1863
              ||| ||  ||||| ||||||||||||||| || ||| |||||||||||||
Db   1508 CCATCGAGGCCGGTCACCCCGCGAGGTTCACGGCACAGCCGCCGCTCCTCAAGCTGACCAAGG 1567
```

```
Qy   1864 ACAAGTTCAAGATCGAGTTCGGCTCCGACACCACCTRACACCCTTCAAGGGCGGACCATCG 1923
              |||  |  ||| |||||||||||||||||||||  ||| ||||||||||| ||| |||
Db   1568 GCACGCAACAAGCT--------------------CACGATCTCGGCGTTCGGGGCGGGATGA 1603

Qy   1924 TCCCCGGCCACGCGTCCTCCTCCCGCGTGGCGGGGGTTGCAAGGTCATTGACGACGCAGGCCGAAA 1983
              |||||| |||||| ||||  ||||||||||||||||||||||||||| || |||||||||
Db   1604 GTGCCACCCGCGCTCTCCCTGGAGCTGGCGCTGGTTGACGGCGGCGACCAGGCGGACGA 1663

Qy   1984 TCGAAAAGTCCGTGCGCCCTCGCCGAAGGAGCAGACGGTCATCATCTGCGCGGGCGCTTA 2043
              |||  ||| ||||||| |||||||||| |||||  |||||||| ||||||||||||||
Db   1664 TCCGCGAAGGCCGTGGGAGTCGGCGCGAAGGGCGTACGGGCGGGGCGTACGGCGGACGA 1715

Qy   2044 ACGGCGACTGGGAGGACACCGGAGGGCGGCGACGCGGCCGAGCAGGTCGAGCTGTCGTCG 2103
              |||||||||||| ||||||||||||||||||||||||||||| |||||||||| ||||
Db   1716 -CTACGACGACGGCCTACGGTGTCGCGGACGGCCAACCCGAACCGGTGGTCGTCGTCGCC 1774

Qy   2104 ACCAGCTCATTGCCGACGTGCCGCGCCGGAGGGCGGCGGCCGGCGCATGAAGCTCCCCGGC 2163
              |||  ||| |  ||||||||||||||| ||||||| | ||||||||||||||||||||||
Db   1775 ACAAGCTGATTCTGGCGTGCGGCGGACGCCAACCCGAACCGGTGGTCTCCATCTGG 1834

Qy   2164 GCACCCCCGGAGGAGATGCCCTGGCTCGAGCGCCACGCCCCGCGCCCGTCATTCGAGTACG 2223
              |||| ||||||||||| |||||||||||| |||||||||||||||||| |||| |||| ||
Db   1835 GTTCGTCGGTGTCGATCCGTCGGTCTGTCGGCCAGCCGCCCGTTCAACACG 1894

Qy   2224 GCGGCAACGAGAACGGGCGCAACTCCATTGCCGACGTCGTCTTTGGGGACTACACCCTCGG 2283
              ||||||||||| |||||||||||||||||  |||||| ||||  |||||||||||||| ||
Db   1895 CGGGCCAGGCGGGCGGCCGAGGGCCACCGCGTCGCGGCTCGTCTCTACGGTCAACCGGGACG 1954

Qy   2284 GCAAGGCTGTCCCTCAGCTTCCC---------AAGCGCTGCAGACAAACCCGCGTTTC 2334
              |||  | ||||||||||||||              || |||||||||||||| |||
Db   1955 GCAAGCTCACGCAGCAGCTTCCCGGCGCCGAGGGCCTGTCGGTCGAGAACCAGCGTCGTGA 2014

Qy   2335 TCAACTTCCGCACCCAGCAGGCGGCGGCGCTGTACGGCGGACGGCTTACGCGGTACA 2394
              |||  || ||||||||||||||||| |||||||| |||||||||||||||||||||
Db   2015 CAAGCTACCCGGCGTCGACGACGGTTACCGCAGGCCGCGCGCCCGGCGGACCGGTACC 2074

Qy   2395 GGTACTACGAGTTGCCGACGACAAGGACGTCAAATTCTCCCTTTGGGCCACGGGCGAG 2454
              ||| || | | ||  ||||||||||||||||||||||||||| |||||||||||
Db   2075 GCTGGTTCGGACAAGGACAAGGTCAAGACCGCGCTGTTCCCCGTTCCGGTGTCGTACA 2134

Qy   2455 CCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGAGGGTGT 2511
              |||  | | ||| | |||||| ||| |||||||||||||       || |||| ||||
Db   2135 CCTGCGTTCACGCAGGAGACGCGCCCCCACCTGCGTCGTAGGTCGTGGGGTCTCAGTCA 2194

Qy   2512 CCCTCTCCGTGAAGAACACACCGGCGTCCGTGTGCGCGGGCCGGCCCAGCGTCTCAGTCA 2571
              |||||||| ||||||||||||||||| ||||||||||||||||||||||||| |||||
Db   2195 CGGTCACGGTCCCGCAACAGGCGGAAGCGGCGACGTGTCCCAGGCGGGTCATGGGGTACCTCG 2254

Qy   2572 AGCCCTCCAAGGCGGCCAAGATTAACGGCCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGG 2631
              ||| ||| ||||||||||  |  ||| ||||  |||||||||| ||| ||  ||||||
Db   2255 GTGCCAGCCCGAACGTGAACGGCTCCCGGACGTGAGGAGAGCTCGTGACACGGAGAAGG 2314

Qy   2632 TCGAAACTGCAGCCCGGCGGGGAGACGAAGGCGGTGAC 2665
              |||||||||| ||||||||||| |||||| |||||||
Db   2315 TCTCGCTCGCGCCGGGCGGGGCGGAGGCGGAAGAGACGGTGAC 2348
```

```
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 23
; LENGTH: 2430
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-23

Query Match        47.6%;  Score 175.2;  DB 9;  Length 2430;
Best Local Similarity  5.9%;  Pred. No. 2.4e-33;
Matches 1016;  Conservative 0;  Mismatches 968;  Indels 150;  Gaps 11;
```

```
Qy   544  CGCTCGGTTCCACATTCAACCAAAACTCTGCTGCTGAAGAGGCAGGTAAGATGATGGGCAAAG  603
Db   353  CCCTGGCCAGCACCTTCGACGACACCATGGCCGACAAGTCATGGGCCGCG  412

Qy   604  AGGGCCATCGGTCTAAGAGTGCGGCATGTGATCGTCTCGGGACACCGCTCCC  663
Db   413  ACGGTCGCGCCGCTCAACCAGGACATGGTCGCTCTGGGCCCATGATCACGGGTGC  472

Qy   664  CTCTCGGTGGACGTGGCGCTTCGAGTCGGATTGGTGGTGAGGAGATCCGTTCCTGCGGCGTTGGGAG  723
Db   473  CGCACGGCGGCGGGAACTACGAGACCTTCAGCGAGGACCCCCTTGGTCGTCGTCCTCGCGCACCG  532

Qy   724  CTGCGGCCTCTCATCCGCGGGCATTCAGAGCACTGGAGTGGCAGCTACGATCAAGCACTTTT  783
Db   533  CGGTCGCCCAGATCAAGGGGCATCCAAGGGTTGCGACACCGCCAAGCACTTCG  592

Qy   784  TGTGCAATGATCAGGAGGACGAGGGCCATGAGTGGTGTCACGAGAGCATCGTCACGGAGGGGGTC  843
Db   593  CGGCCAACAACCAGGAGAGAAACACGCGCTTCTCCGTGAAACGAGAGCAGC  652

Qy   844  TCCGTGAAAATCTACCGCACTCCCGTTCCAGAATTGCTTGTGTGGGGAGAGACTGGTACGGCA  903
Db   653  TCCGCGGAGATCGAGTTCCCGGCGGTTCGAG---GGCTTCCTCCAAGCCGCGGGCGCCTCCT  709

Qy   904  TCATGACGGCGGTACAATGGCATCAATGGCGTGTCGTGCTCAGCGAGAGGGGGTC  963
Db   710  TCATGTGTGCTCTACAACGGGCCTCAACGGAGGAAGCGTCCTCCGCGGCCGGGCTCTA  769

Qy   964  ATGGGATGGCTTGGAAAGGAATGGGGTTTGGGGATGGCTAATCATGAGCGACTGGTACGGCCA  1023
Db   770  ACAACGTCGTGCCCACGGCCGCAGTGGGGGCGCTTCCCAGGGCTGGGTGATGTCCGGACTGGGCTCGCCA  829

Qy  1024  CATACAGTACCACAGAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCCGGACCTCGAC  1083
Db   830  C---CCCGGGCACCGACGCCATCACCAAGGGCCTCGACCGACGAGATG  873

Qy  1084  GCTTCCGAGGAGGAAAACACTCAAGTTCAACGTCTCCAACGGAGAAAGCCCTTTATCCACGTCA  1143
Db   874  ------GGCGTCGAGCGTCCCCGGCGACGTCGGGCGGGCGCTGGGTGATGTCCGGACTGGGCTCGCCA  925

Qy  1144  TTGACCAGAGAGGGCTAGGGAAGTTCTTCGTCAAGAAGTGTGTGCTTGCCCTTCGCGGAGTGA  1203
Db   926  AGTTCTTCTGGGCGA---GGCGCTGAAGAGACGGCCGTGCACGGGTGCCCCGGAGGCGG  982

Qy  1204  CGGAGAACGGCCCCGAGAGAGAGACTGTCTCAACAACAACCCCGAAACGGAGCAGCTTCCCGGA  1263
Db   983  CCGGGCGGTGCCGGAGAGACGGAGAAGTTCGGGTCTCGCCTCCTCG  1042

Qy  1264  AGGTTGGGCAACGAGCGAGGGCGCATCGTGCTGCTGAAGAGAACGAGTTCTGCCCTTGAGCA  1323
Db  1043  CCACTCCGGCGGCGCGGCCGGGCCCCGAGCGCCGGACAAGCGGGGTGCCCGCAAGG  1102

Qy  1324  AGAAGAAGAAGACCGCTGATTGTCGCGCCCAAACGCCAACGCGACCATACCACGGCCGGAG  1383
Db  1103  TCGCCGGAGAACGACGGCGCGGTGCTGCTGCTGCGCGCTCGCGCGGTG  1162

Qy  1384  GCTTGTGCCCGCCACTCAGGGCCTCACTCCTCTTTGACGGCGTCAGCAAGCAGC  1443
Db  1163  ACGCCGCGGCAAGAGCATCGCGCCGGTCAATCGGCCGCCGACGCTCGAC--CCCAAGGTCACCGGC  1221

Qy  1444  TCGAGACGCCGCCGCCATCGTACACCGTCGGGGCGTTCCTCCCATTCTAGGCG  1503

Db  1222  CTGGTACCCGCTAACCGGCGCCCACGTCGTCCGGGACTCGGCCCACTCGACACCATCAAGGCC  1281

Qy  1504  AGCAGTGCCTCAGCGCCGGACCCCGACCGGCCCTCCGGGGCATGCCCTGGAGGGCACGT  1563
Db  1282  CGCCGCGGGGTGCCGGACGGTGCGACGTACGAGAGGGGAGGAGACCTTCGGGACGCAG  1341

Qy  1564  CTGGTACCCCTTAACCGCCGCACCACATTGACGAGCTTCTTCACCAAGACGCACC  1623
Db  1342  ATCCCGGCCGGGGGGGAACCTCAGC------CCGGCGTTCAACC  1375

Qy  1624  TGGTGGACTACTACCACCCCCAAAGCGGCGGCAGACACGTGGTACGCGGCAGGGGCACGT  1683
Db  1376  AGGGCCACCAGCTCGAGCGGCGCAAGGGCGGGCTGTACGACGGCCAAGG  1435

Qy  1684  ACACCGCCGACGAGGACTTGCACCTTAGCAGGCTCGGGCTTCGTGCGGCGCAACCATCG  1743
Db  1436  CCGCCCGACGGCGGAGTACCGCCATCCGGGGTCCGCCACCGGGTGGTTACCGCACGGTGCAG-  1494

Qy  1744  CGTACGTTAGACGACCAGCTCGTCGTCGACAACGCCACCAGCAGGTCCCCGGGGATGCCT  1803
Db  1495  ------GCACGGCACAAGCT--  1507

Qy  1804  TCTTCGGGCTTCCGGCCGCCACCCGGGGAGGACGGCGGCGATCAAATCTCTGGTCAAAGGGCACGT  1863
Db  1508  CCATCGAGGGCCGGTCAGGTCTTCAGCGGCAAGGTTGAGCAGACCCGGCTCCCTTCAAAGCTGACCAGG  1567

Qy  1864  ACAAGTTCAAAGATCGAGTTCGGGCTCCGCCACCCTTCAACCCGGGGGACAATCG  1923
Db  1568  GCACGCACAAGCT------CACGATCTCGGGCTTGGGCATGA  1603

Qy  1924  TCCCGGGCCACGGCTCCCTCGCGTCGGGCGAGGGAACGGACGGCCGGATCAAATCTCTGGTCAAAGGGCACAGT  1983
Db  1604  GTGCCACCCCGGCTCTCCTCGGAGTCTGGGGCTGACGCCGGGCCCTACGGCGACGA  1663

Qy  1984  TCGAAAAGTTCCGTCGCGCCCCTCGCCAAGGACTGGACGGGGGGCCTTA  2043
Db  1664  TCCGGAAGGCCGTGAGTCGGGCGGCGGAAGGGGCACGGGTCGGTC------  1715

Qy  2044  ACGCCGACTGGGAGGACGCGAGGGCGCGGCCCGACCGGCGGACGATGAAGCTCCCCGGGTGCTGG  2103
Db  1716  -CTACGACGACGCCCACCGGGCGTCAGCAGTGCCGAACCGGTCCTCGATCACACGGGGGTTACGG  1774

Qy  2104  ACCAGCTCATTGCCGACGTTGGCCGCCCGGGAGAACCCAAACACCGTCTGCGTCGTCGGGACGG  2163
Db  1775  ACAAGCTGATTCTCGGCTGTCGCGCGGACGGCCAACCCGACACGATGTGGTCCTTCAACACACG  1834

Qy  2164  GCACCCCCGAGGAGAGATGCCCTGGCCTTCGACGCAGGACAAACCCCGCGTTTC  2223
Db  1835  GTTCGTCGGTGCGTGGCTGGGTGATGTCCGGCCGTCCAAGCGCCGACATGTGGTTACC  1894

Qy  2224  GCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGGACTACAACCCCTCGG  2283
Db  1895  CGGGGCCAGGGGCCGGGCGGGACCACCGCCGGGTTCAACGCCTTCACCCCGAACGACG  1954

Qy  2284  GCAAGCGTGTCCCTCAGCCTTCCCC--------AAGCGCGCTGCAGGAGACAAACCCGCGTTTC  2334
Db  1955  GCAAGCTCACCCAGAGCTTCCCGGCGCACCACGGGTCGCCGGAGGACACGCGGGGTACC  2014

Qy  2335  TCAACTTCCGCACCCAGCGCGGGCGCGCGTCGTCGCCGCGCTGTACGGGCCGGGTACA  2394
Db  2015  CAAGCTACCCAGAGCTTCGACAACGCGACAACGACGTACGCGTCGGGGGACCG  2074

Qy  2395  GGTACTACGAGTTGCCGACAAGGACGTCAAATTCCCGTTTGGCCGCCGTGTCCTTACA  2454
Db  2075  GCTGGTTCGACAAGAGGAAGCTCAAGCGCGTTCCCGTTGGCCGGCGTCACCGGAGGAC---  2134

Qy  2455  CCACTTTGCCTTCCAAATCTCTCGGTGTCTCACAAGGAC---GGCAAGCTGAGGCGTGT  2511
Db  2135  CCTCGTTCACCGACGAGAGCGCCCCGACCGTGTACGTCGGGGTGGTTCTGGAAGGTCA  2194

Qy  2512  CCCTCTCCGTGAAAGAACACACCCGGCGTCCGTGCCCAGCGTCTCACGTCA  2571
```

```
Db  2195 CGGTCACGGTCCGCAACGCAGCGCGCGGAAGCGCGCCGGAGCGAGGTCGTCCCAGGCGGTACCTCG
Qy  2572 AGGCCCCTCCAACGCGCCAAGATTAACGCGCCCGTCAAGGAGCTCAAAGGGGCTTGCAAAGG 2631
Db  2255 GTGCCAGCCCGAACGTGACGGCGCTCCCGCGAGCGAAGAAGAAGCTCGTGGGGTACACGAAGG 2314
Qy  2632 TCGAACTGCAGCCCGGACGCGGGGGGAGACGAAGGCCGGTGAC 2665
Db  2315 TCTCGCTCGCCGCGGGGCGGAGCGCGAAGACGGTGAC 2348
```

RESULT 3
US-09-861-289-23
Sequence 23, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
  APPLICANT: Sherman, D.H.
  APPLICANT: Liu, H.
  APPLICANT: Xue, Y.
  APPLICANT: Zhao, L.
  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
  FILE REFERENCE: 600.438US1
  CURRENT APPLICATION NUMBER: US/09/861,289
  CURRENT FILING DATE: 2001-05-18
  PRIOR APPLICATION NUMBER: 09/105,537
  PRIOR FILING DATE: 1998-06-26
  NUMBER OF SEQ ID NOS: 43
  SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 23
  LENGTH: 2430
  TYPE: DNA
  ORGANISM: Streptomyces venezuelae

Query Match 5.9%; Best Local Similarity 47.6%;
Matches 1016; Conservative 0; Mismatches 968; Indels 150; Gaps 11;
Score 175.2; DB 10; Length 2430; Pred. No. 2.4e-33;

```
Qy   544 CGCTCGGTTCCACATTCAACCAAACTCTGCTGAAGAGGCAGGTAAGATGATGGGCAAAG 603
Db   353 CCCTGGCCAGCACCTTCGACGACACAGCTACGGCCAAGGTCATGGGCCGCG 412

Qy   604 AGGCCCATCGCTAAGAGTGCGCATGATGATCTCGGCGCCGACTATCAACATGCAACGGCTCC 663
Db   413 ACGGTCGGCGCGTCAACCAGGACACATGGTCGTCTGGGCCCGATGATGAACACAATCCGGGTGC 472

Qy   664 CTCTCGCGTGGGACGTGGGCTTCGATTGGTGGTCAACGGGATCCGTTCGTTCTCTGGCGGCTTGGGAG 723
Db   473 CCCACGCGCGGCGGGAACTACGAGACCTTCAGCGAGGACACCCCTGGTGCTTCCTCGCGCCACCG 532

Qy   724 CTGCGGCTCTCATCCGCGGGGCATTCAGACGCACTGGAGTGCAGAGGCTACGGATCAAGGACTTTT 783
Db   533 CGGTCGCCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCCAAGGCCGCGGGCCTCCT 592

Qy   784 TGTGCAATGATCAGGAGGACAGAGGGCCGCATGAGTCGTCGAGAGCATCGTCACGGAGGGGCTC 843
Db   593 CGGCCAACACCAGGAGAAACACCGCTTCTCCGTGAAGCCAATGTCGACGAGCGAGGAGC 652

Qy   844 TCCGTGAAATCTTACGCGACTCCGTTCCCAGATTGCTGCCGAGACTCCCAGCCGTGCGT 903
Db   653 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCCAAGGCCGCGGGCCTCCT 709

Qy   904 TCATGACGGCGGTACAATGGGCATCAATGGGGTGTCGTGCAGCGAGAAACCCTAAATATATCTTG 963
Db   710 TCATGTGTTGGCCTACAACGGCCTCAAGCGGCCGTTCCAGGCTGGTGAGTCCGACTGGCTCGCCA 769

Qy   964 ATGGGATGCTTCGGAAAGGAAATGGGGTTTGGGATGGCCTAATCATTGAGACGGTACGGCA 1023
Db   770 ACAAGGTGCTGCGGCAACGCAGTGTGGGCGCTTCCAGGCGTGGTGGATCGCGGGAAGGCCTCCTCA 829

Qy  1024 CATACAGTACCACAGAAGCCGTTGTGGCAGGCGTCGACCTCGAGATGCCCGGACCTCCAC 1083
Db   830 C--CCCGGGCGACCGACGCCCATCACCAAGGGCCTCGACCAGAGAGATG----------- 873
```

```
Qy  1084 GCTTCCGAGGAGAAACACTCAAGTTCAAACGTCTCCAACGGAAGGCCCTTTATCCAACGTCA 1143
Db   874 ------------GGCGTCGAGCGTCCCCGACGGAAGGTCTCAAGAAGTGTCGCCTCGGGAGTGA 925

Qy  1144 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGCAAGAAGTGTGCCTCGGCGCCGGAGTGA 1203
Db   926 AGTTCTTCGGCGA---GGCGTGAGGAGCTCCCCGGCAAGGTCCCGACGGGGTCCCGGACGGG 982

Qy  1204 CGGAGAACGGCGCCCGAGAGACGATGTCAACACACCCCGAAACGGCAGCGTCTCCTCCGGA 1263
Db   983 CCGTGACGGCGGTCGGCGGAGCGGGGAGATCGTCGGCGGGGATCGTGCTCGGCCTTCGCTCG 1042

Qy  1264 AGGTTGGCAACGAGGAGCGCATTCGTGCTCGCTGAAGAACGAACAACGGTTCTGCCCCTTGACGA 1323
Db  1043 CCACTCCGGCGCCGCGGGCGGCGCCGAGCGCGACGACAGGGCCGGTGCCCAGGCGGGTCCCGCAAGG 1102

Qy  1324 AGAAGAAGAAGACGCTAGTGTCTGCGCCCCAACGCGCCAAAGCAGGCGCACATACCACGGCGGAG 1383
Db  1103 TCGCCGAGAACGGGCGCGGGTGCCTCGGGTGTGTCGCCAACGCCGCAGGCGGCTGCGCGTCCGGTG 1162

Qy  1384 GCTCTCCCGCACTCAGGGCCTACTTACCGCACCCTTTGACGGCCCTCAGCGAAGCAGC 1443
Db  1163 ACGGCCGGCGAAGAGCAATCGCGGTCATCCGCCCGACGGGGCTCCATCGGCGTCGAC-CCCAAGGTCACCGGC 1221

Qy  1444 TCGAGACGCGCCCATCGTACACCGTCGGGGCCTACACCACCGTTCCTCCCATTCTAGGCG 1503
Db  1222 CTGGGCGACGCGGCCCACGCCCCAGTCCGGCGCCGGCGCGTCGGGCGTCGGCGCCGGTGTG 1281

Qy  1504 AGCAGTGCCTCACGCCCGACGGGGCGCGGGCCGCTCCGGGCATGCGGCGCTCGCGGCGTG 1563
Db  1282 CGCGCGGGGTCGGGGTGCGACGGTGACGTACGAGAGGACTTCGGGACGAG 1341

Qy  1564 CTGGTACCCTACCGCGCAGCACAGCACATTGACGAGCTCTTCTCCACCAAGCACGACGATCCACC 1623
Db  1342 ATCCGGGCGGGGGGGAACCTCAGC---------------CCGGCGTTCAACC 1375

Qy  1624 TGGTGGACTACTACCACCCCAAGGGCGGCAGACACGTGGTACGGCCGACGACATGGAGAGGGCACGT 1683
Db  1376 AGGGCACCGCCCCACGGGAAGCGGGCGGCGTTCATCAACGAGGGCGGGGAGGGCACGTGGC 1435

Qy  1684 ACACGCGCGACGAGGACTGCACCACTACGGACGCTCGGCCGCTCGGGGGCTGCGGGGCAAAG 1743
Db  1436 CCGGCGACCGGCAGTACCGCATCGGCGTCCGTGCCGACCGGGTGTTACGGCCGGCCGCAG- 1494

Qy  1744 CGTACGTAGACGACCAGCTCGTGTCGTCGGACGCCACCAGCAGGTCCCCGGGGATGGCCT 1803
Db  1495 ---------------------CTCGGCAGCCCACA 1507

Qy  1804 TCTTCGGCTCCGGCAGCAGCACCCGCGAGGACAGGACGGCGCCGCATCAATCTCGTCAAGGACCAACACGT 1863
Db  1508 CCATCGAGGCCGGTCAGGTCTTACGGCCGGCCGCAAGGTGAGCAGGTCGCAAGCAGGCCGAAA 1567

Qy  1864 ACAAGTTCAAGATCGAGTTCGGCTCCGCACCACCTACACCCTCAAGGGCGGACACCATCG 1923
Db  1568 GCACGCACAAGCT-------------CACGATCTCGGCGCTTCGGGATGA 1603

Qy  1924 TCCCCGGCCACGGCTCCCTCCGGCGTCGGCGGCTGCCAAGGTCATTGACGACCAGGCCGAA 1983
Db  1604 GTGCCGCACCCCGCTCCCCTGAGCGTGGCGTTGACGGCGCCGAAGGTCGGCGGGACGAGACGA 1663

Qy  1984 TCGAAAAGTTCGGTCGCGCCTCGCCACCACCTACACCCTCGCGGGCGCCCTTA 2043
Db  1664 TCGCGAAGGCCGTGAGTCGGCGGGAAGGCCCGTCACGGCGTCGTTCTCTCA 1715

Qy  2044 ACGCCGACTGGGAGACCGCGGACGCGCCGAGCGATGAAGAGCTCCCCGGCTGCTGG 2103
Db  1716 -CTACGACGACGACGGACGACGAGGGCGTCGACCGGTCGAACCCTGTCGCGTGCGTCCTGCGAGG 1774

Qy  2104 CATACGAGTACCACAGAAGCCGTTGTGGCAGGCGTCGACCTCGAGATGCCCGGACCTCCAC 2163
Db  1775 ACAAGCTGATCTCGGCGCTGTCGGCGGCCAACCCAACACGATCGTGGCGTCCTGCGAGACGG 1834
```

RESULT 4

US-09-988-384B-3

Sequence 3, Application US20030073824A1

Publication No. US20030073824A1

GENERAL INFORMATION:

APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.536US1
CURRENT APPLICATION NUMBER: US/09/988,384B
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: PCT/US99/14398
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: US 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 53

SEQ ID NO 3
LENGTH: 12441
TYPE: DNA
ORGANISM: Streptomyces venezuelae

US-09-988-384B-3

Query Match   5.9%;   Score 175.2;   DB 9;   Length 12441;
Best Local Similarity 47.6%;   Pred. No. 5.2e-33;
Matches 1016;   Conservative 0;   Mismatches 968;   Indels 150;   Gaps 11;

```
Qy   2164 GCAACCCCCGAGGAGGAGGAGGATGCCCTTGGCCTGCGACGCCACGCCCGCGTCATCCAGGCCCTGGTACG 2223
          |||||| |||   ||| |||||     | ||||  || |||||| | || ||||| || |||  ||||
Db   1835 GTTCGTCGGTCGGTCTGATGCGTGGTCGTCAAGACACCCGGCGGGTCCTGGACATGTGGTACC 1894

Qy   2224 GCGGCAACGAGACGGGGCGAACTCCATTGCCGACGTCGTCTTTGGGGACTACAACCCCTCGG 2283
          |||   || |||||||||||||||||||||  |||||||||||||||||||||||||
Db   1895 CGGGCCAGGCGGGCGGGGGCCGAGCGCCACGCGCGCCGCCTCTACGGTGACGTCAACCCGAGCG 1954

Qy   2284 GCAAGCGTGTCCCTCAGCTTCCC-----AAGCGTCGCAGGACAACCCCGCGGTTTC 2334
          |||||||||| ||||      |       ||||| ||  |  ||||||||||| ||
Db   1955 GCAAGCTCACGCAGGAGCTTCCCGGCGCCGAGCAACACGCGGTCGCGCGGGACCGA 2014

Qy   2335 TCAACTTCCGCACCGAGGCCCGGGCCACGCTTGTACGGCGGAGGAGACGTTCTACGTCGGGGTACA 2394
          |||||||||||||||||   |||  |||||  ||||||||||  ||||||| ||||||||||
Db   2015 CAAGCTACCGGGCGTGCGACAACCACGAGAACTAGCGAGAGGGCCATCCAGTCGGGTACC 2074

Qy   2395 GGTACTACGAGTTTGCCGACAAGGAGAGTCAATTTCCCCTTTGGCCACGGCCCTGTCCTACA 2454
          ||||   ||  ||  ||| ||||  || |||||| |||||||||| || |||||||||||||
Db   2075 GCTGGTTGACAAGGAGAAGGTCAAGCCGCTGTTCCGTTTCCCGTTCGGGCTGGACACGGCCCTGTCCTACA 2134

Qy   2455 CCACTTTTGCCTTTTCCAAATCTCCGTGTCTTCACAAGGAC---GGCAAGCTGAGCGTGT 2511
          ||||||  ||| ||  ||| ||||  || ||||| |||||     ||||| || ||| ||
Db   2135 CCTCGTTCAACGACGCCGTCGTTCCGTACGGTCGCACGGGTGGGGTGGGGTCGGGGTCA 2194

Qy   2512 CCCTCCGTGAAGAACAACCACCGGTCTCGGTGCCCGGCGGGCCACAGTTGGGCCCAGCTCACGTCA 2571
          |||  || | ||     ||  |||| ||| ||| ||| |||||||| |||||  ||| || ||||
Db   2195 CGGTCACGGGTTCCGGCAACGAGGGCGTCGCCGGCGGCAGCGGTCGTTCCACGGCGTACCTCG 2254

Qy   2572 AGCCCCTCCAAGCGGGCCAAGAGATTAACCGCCCCGTCAAGGAGGTCAAGCCGGCTTCGCAAAGG 2631
          ||| || |||||||||||||| |||  ||| |||||| |||||| || ||||| ||||| || ||
Db   2255 GTGCCAGCCCGAACGTAGCGGCTCCGCCAAGGGTCGCGCGAGAAGAACGCTCGTTCGGGCTTACACGAAGG 2314

Qy   2632 TCGAACTGAGCCCGGACCGGGAGAGACCGAGACGAAAGGGCCGGTGAC 2655
          |||  || || |||||    |||  ||  |||| ||| ||||||
Db   2315 TCTCGCCTCCGCCGGCGGCGGCGAGGCGAAGACGGTGAC 2348
```

```
Qy    664 CTCTCGCGTTGGGACGTGGCGTTCGAGTTCGAGTTGGTTGAGAGATCCGTTCTTCTTGGGGAGATCGAGG 723
          ||  || ||| |||| ||||   || ||||| |||| ||||| ||   |||||||||  |||||| ||||
Db   4097 CGCACGGCGGCGGCGGGACTACGAGACCTTCAGCGAGGAGGACCCCCTGGTCTCCTGCGGGGCGGCTGGGAG 4156

Qy    724 CTGCGGGCTTCATCCGGGCGGGCATTCAGAGCACTTCAGCGAGGAGGTCAGCGCTACGGATCAAAGCACTTTT 783
          ||||   || | ||  ||| ||| || ||  |||| ||||  || |  |||||||| ||||  ||||||| ||
Db   4157 CGGTTCGCCCAGATCAAGGGCATCCAGGTGCGGGGTCTGATTGACCAGGGCCAAGCACGTCGACTTCG 4216

Qy    844 TCCGTGAAATCTACGGCACTCCCGTTCACGACCTTCTGCGTGTGGCTGGACCGCTCGGTGCGGT 903
          |||| |||| ||| | ||  |||| |||  ||| ||  ||||| |||| ||  |||| || ||||
Db   4277 TCCCGGAGATCGAGTTCCCGAGCGCACGAGAGGGCGAACGCGCGGGGGCGGTCGGCCA 4333

Qy    904 TCATGACGGCGTACAGGTGCATCAATGATGCGTGTCGTCGCAGCGGAGCAGAACCTAAATATCTTG 963
          || ||||| |  ||  |||||  ||||   | | || |||  || ||||| |||   ||| |  ||
Db   4334 TCATGTGTCCTACAACGGCCCTCAAACGGCGAAGCGCCTGCGCAACGGCGAGCTCCTCA 4393

Qy    964 ATGGGAATGGCTTCGAAAGGAAATGGGGTTGGGATGGCCTAAATCATGAGCGGACTTACGGACG 1023
          ||  | ||| |  || ||| ||   || ||||  ||||| |  | ||||| ||| ||||| ||||
Db   4394 ACAACGTGCTGCCGGCCAGTGGGGCTTCCAGGGCTGGGTGATGTCCGACTGGCTCGCGCA 4453

Qy   1024 CATACAGTACCACACAGAAGCCGTGTGTGGGCAGGGCCTTCGACCTCGAGGAATGGCCGGGGACCTCCAC 1083
          |||  | ||  |   |||  |  ||||| ||||| ||||||  ||| ||||   ||||||  || ||||
Db   4454 C---CCCGGGCGCACGGGCCGCCATCACCAACCAGGGCCTTCGACCAGGAGATG 4497

Qy   1084 GCTTCCGAGGAGAAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCTTTATCCACGTCA 1143
          |||   || || | || | |  || | ||| ||  |||| | ||  | || || || ||
Db   4498 -------GGGTCGAGCTCCCCGGCGGTCGAGGAGGGGCCGGGCCGCGGCCA 4549

Qy   1144 TTGACCAGAGGGCGTAGGGAGAGTCTTTCAGTTTCGTCGTCAAGAGAAGTGTTGCTCGGAGTGA 1203
          |||||| ||||| || |||   | |||| ||  |||||||||||||||| |  | |  ||| ||
Db   4550 AGTTCTTCGGGCGA---GGGCGCTGAAGGACGGTCCTGAACGGCGACTTCGCCGGGGG 4606

Qy   1204 CGGAGAACGGCGCCCCGAGACGAACTGTCAACAACACCCCCGAAAACGGCAGCTCTCTTCCGA 1263
          | | ||||| | || |||   || |||| ||||| |||||||  ||| || |  ||| | |||
Db   4607 CCGTGACGCGGTCGCGGGACGGCGGATCGTCGGCCCAGATGGACAGAAGTTCGGTGCTCCCTCG 4666

Qy   1264 AGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACACGAGGAACACGTTCTGCCCTTGACA 1323
          | | ||| ||| || |||||||| || ||  ||| | || || ||| |||||  ||||||
Db   4667 CCACTCCGGCGCCCGGGGGCGGGCGGACGGCGGAACGGCGTCCCGGCAAGG 4726

Qy   1324 AGAAGAAGAAAGAACGCTGATTGTCGGCCCCAAACGCAAGCAGGCCACAATACCACGCGGAG 1383
          | |  || ||| |||||  |  || ||||    |  ||| |||||||| || | |||| ||
Db   4727 TCCGCCGAGAACGGGCGCGGTGGTCCCTGGCTCTCCGCGGTGCGGCCCTCGCCGGTG 4786

Qy   1384 GCTCTGCGCGCACTCAGGGCCTACTACGCAGTCACTCCCTTTGACGGCGGCCTCAGCAAGCAGC 1443
          ||||| ||||  ||  |||| || |||||    |||||| ||||||| || || |||| |||
Db   4787 AGCCCGCGCCAAGAGCACTGCGGGTCATGCCGACGTCGCGAC-CCCAAGGTCACCGGC 4845

Qy   1444 TCGAGACGACGCGCCGCCATCGTACACCGTCGGGGCCGCCTACACCACCGTTCCCCCATTCTAGGCG 1503
          | ||| ||| |  |||| | |  | || |||||| || ||| |  ||  ||  |||  || ||| |
Db   4846 CTGGGCACCGCCGCCACGTCGTCCGGACTCGGCGCCGCCACTCGACACCATCAAGGCC 4905

Qy   1504 AGCAGTGCCTCACGGCGACGGCCGACGCTCCGGGGGTTCTTCAACGAGGCCC 1563
          | ||||| | ||  |  ||||||| |||  |||||||  || ||||  ||||
Db   4906 CGCGCGGGGTGCCGGGCGTGGCAGCGGTCTACGGACGGGACCTTCGGACACGCAG 4965

Qy   1564 CTGGTACCCCTAACCGCCGCACATTGACGAGGCTCTTCTTCACCAAGAGCTCCGCACC 1623
          || |  || ||  | |||  ||| |||||||| | || ||| || ||    ||  | |
Db   4966 ATCCCGGGCGGGGAGAACCTCAGC---------CCGCGCGTTCAACC 4999
```

```
Qy    544 CGGTCGGTTCCACATTCAACCAAAACTCTGCTCGGAAGAGGCAGGTAAGATGATGGGGCAAAG 603
          ||| ||  || |  | |   ||| |  |  |||  ||| || | |||   ||| || || |||
Db   3977 CCCTGGCCAGCCGCACCTTCGACGACCAGCTACGGCGAAGGTCATCGGGCCGCG 4036

Qy    604 AGGCCATCGGCTAAGAGTCGCGCATGTGACCACCACACTGGCCGAAGGTTCCGGGCCTCCC 663
          |   || || | ||||| | || | |||  | |  || ||||  | ||||| ||||| ||
Db   4037 ACGGTGCGGCGGCTCAACCAGGACACATTCCGGGTGC 4096
```

```
Qy   5000 AGGGCCACCAGCTGAGCGCGGCCAAGGCGGGGCGACGGGGGACGCTGACCGTG 5059
          ||| |   ||||  ||||| |||    ||||||  ||| || |||| || |||
Db   1624 TGGTGGACTACTACCACCCCAAGGCCGGCGAGACACGTGGTACGGGGACGGT 1683

Qy   5060 CCGCCGAGGGGCGAGTACCGGCCGTCCGTGGCGTTCGACGGCCTCCAG- 5118
          |||  ||||| |  | || |||  |  |||||  |  ||||||||||||
Db   1684 ACACCGGCACCGCGAGTACGGGACGGTGGGCGTTGGGCCTCCAGGCGGG 1743

Qy   1744 CGTCAGTAGACGACCAGCTCGTCGTTCGACAACGGCACCAAGCAGGTCCCGGGCGATGCCT 1803
          || ||  || |  ||||  ||| ||| |||||||||||||  |||| ||| ||| | ||
Db   4037 ACGGTGCGGCGGCTCAACCAGGACACATTCCGGGTGC 4096
```

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
. NUMBER OF SEQ ID NOS: 43
. SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-3

Query Match 5.9%; Score 175.2; DB 9; Length 13613;
Best Local Similarity 47.6%; Pred. No. 5.4e-33;
Matches 1016; Conservative 0; Mismatches 968; Indels 150; Gaps 11;

```
Qy    544 CGCTCGGTTCCACATTCAACCAAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 603
          ||||  |||  |||| ||||       ||||||||| ||||| ||||||||| ||||| ||
Db   4664 CCCTGGCCAGCCACCTTCGACGACACCATGGCCGACAGCTACGGCAAGGTCATGGGCCCCG 4723

Qy    604 AGGCCATCGCTAAGAGTGCCGCCATGTTGAATCCTCGGCCCGACTATCAACATGCAACGCTCC 663
          ||||| |||  || ||  || |||| |  |||  ||||||||| ||||  ||| || |||||
Db   4724 ACGGTCGCGCCCTCAACCAGGACACATGGTCCTGGGCCGCCATGATGAACAACATCGGGTGC 4783

Qy    664 CTCTCGGTTGGGACGTGGCTTCGAGTCGATGATTGGTGAGGATCCGTTCCTTGGCGGGTTC 723
          |||||   || ||||| |||  ||||||||||||| ||||||| ||||| ||  ||||
Db   4784 CGGCACGGCGGCGGAACTACGAACGCTTCTCCCGTGAAGCCAATGTGCAGACGACGAGACGC 4843

Qy    724 CTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGGTGCAGGTTACAGTTACACACTTTT 783
          ||    || ||  ||||||  |  ||||||||| | ||||| || ||||||||||||||
Db   4844 CGGTCGCCCAGATCAAGGGCATTACAGGGTGCGGGTTCTGATGACCACGGCCAAGCACTTCG 4903

Qy    784 TGTGCAATGATTCAGGAGACAGGGGCCATGATTGGTGCAGACATCGTCACGGAGGCGGCTC 843
          ||||| |||   |||||| ||||| ||||| |||  || |||| ||| ||| ||||| |||
Db   4904 CGGCCAACAACCAGGAGAACAACCGCTTCTCCGTGAACGCCAATGTGCAGACGACGAGACGC 4963

Qy    844 TCCCTGAAATCTACGCACTCCCGTTCCAGATTGCTGGCGAGAGACTCCCAGCCGTGCGCGT 903
          ||||||   ||| ||||| ||| |||||||| ||||| ||||||| ||| |||||||  ||
Db   4964 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCCAAGGCCGGCGGGGCCTCCT 5020

Qy    904 TCATGACGGCGGTACAAATGGCATCAAATGGCGTTGTCGGTGCGGAGAACCCTAAATATCTTG 963
          ||  ||| ||| |||  ||||||  |||||||  |||| ||| ||||| |||  || | |||
Db   5021 TCATGTGTGCCTACAACGGCCTCAAACGGGAAGACGCTGCGGCGCAACGAGCAGAGCTCCTCA 5080

Qy    964 ATGGGATGCGTTCGAAAGGAAATGGGGGTTGGGATGGCCTAAATCATGACGCGACTGGTACGGCA 1023
          |||  |||  || ||  ||  || |||  || |||||| ||  |  |||| || ||  ||||| ||
Db   5081 ACAACGTGCTGCCGACGCAGTTGGGGCTTCCAGGCGTGATGGCGACTGGCTCGCCA 5140

Qy   1024 CATACAGTACCACAGAGACCCGTTGTTGCAGCGCCTGCACCTCGGAGATGCCCGGGACTCCAC 1083
          || || ||| || ||| ||| |||| ||| ||| |||| |  ||||| |  ||| ||||| |
Db   5141 C--CCCGGCCACCGACGCCATCACCAAGGGCCTCGACCAGGCGGACATG------- 5184

Qy   1084 GCTTCCGAGGAGAAAACACTCAAGTTCAACGTCTCCAACGGGAAAGCCCTTTATCCAGCTCA 1143
          |||| ||||||||| || |  |||| |||| | || |||||||| || |  |  | ||| |
Db   5185 ------GGCGTCGAGCTCCCCGGGCCGCGTCCCAACGGGGCCCGAGCGGCCGCCA 5236

Qy   1144 TTGACCAGAGGCGTAGGGAAGTTCTTCAGTTGCTCAAGAAGTGTGCCTCCGGAGTGA 1203
          ||  ||  |||  || |||||  |  |||  || |||  || | || || || ||
Db   5237 AGTTCTTCGGCGA---GGCGCTGAACGACGCGCTCCTGAACGGCGGAGGCCG 5293

Qy   1204 CGGAGAACGGCCCCGAGAGACGACTGTCAACACACCCCGAAAACGGCAGCTCTCTCCGGA 1263
          | || |||||  || ||| ||||  ||||| ||||| ||||| ||||||| ||  |||
Db   5294 CCGTGACGCGGTGTCGGGGAGGGAGTCGTCGGCCAGATGGAGAGAGTTCGGGTCTGCTACCTCG 5353

Qy   1264 AGGTTGGCAACGAGGGCCATCGTGCCTCTGAGAACGACGAACACGTTCTCGCCCTTGAGCA 1323
          || |||| ||| || ||| |||||| |||| |||| || ||| ||  ||||||||  ||
Db   5354 CCACTCCGGCGCGCCCGGGCGGCGGCGGACAAGGCCGGTGTCCGACAAGG 5413

Qy   1324 AGAAGAAGAAGACGCTGATTGTCTGCGCCCCCAACGCCAAGCCACATACCACGGCGGAG 1383
          ||||
Db   5414 TCGCCCGAGAACGACGCGCGGTGCTGCTTCCGCGACAACGAGGGCCTGCGCGTCGCGGTG 5473
```

RESULT 5
US-09-860-846-3
; Sequence 3, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.

```
Db   5119 ----------------------------------CTCGGCAGCCACA 5131

Qy   1804 TCTTCGGCTCCCCACCCGCGAGGAGGACGGGGCCGATCAATCTCGTCAAGGGCAAACAGT 1863
Db   5132 CCATCGAGGCCGGTCAGGTCTTACGCGCAAGTTGAGCAGCCGCAAGCTGACCAAGG 5191

Qy   1864 ACAAGTTCAAGATCAGAGTTCGGCTCCGCACCACTTACACCCTCAAAGGGCGACACCATCG 1923
Db   5192 GCACGACAAGCT---------------------CACGATCTGGGCGTTCGGATGA 5227

Qy   1924 TCCCCGGCCACGGCGTCCCTCCGTCGCGGCGTGCCCAAGGTCATTGACGACCAGGCGCGAAA 1983
Db   5228 GTGGCCACCGCCGGCTCCCCTGGAGCTGGGGTGGACGGCGGCGACCGACCGA 5287

Qy   1984 TCGAAAAGTCGTCGGCCCTTGGCCCAAGGAGCAGCAAGTCATCATCTCGGCGGGGCCTTA 2043
Db   5288 TCGCGAAGGCCGTGGAGTCGGCGCGGGGCGAAGGCGGGTCGTCTTCGC----- 5339

Qy   2044 ACGCGCGACTGGGACGACCGAGGGGCCGCGAGCATGAAGCTCCCCGGGCCGGTGCTGG 2103
Db   5340 -CTACGACGAGCGCACGAGGGCGTGACGTCGCGGCCGACACCGTCTCGGCGGGGTACGCGAGG 5398

Qy   2104 ACCAGCTCATTGCCGACGCGCGGCCGCCGCGGGGCGCCAAAAACACGTCGTCGTCATGCAGACGG 2163
Db   5399 ACAAGCTGATCTCGGCGTGTGCGGGGACGCCCAACCCGGAACCGTGGTTCCTCAAACACG 5458

Qy   2164 GCACCCCGAGGAGATGCCCTGGCTCGACGACCACGCCCGCGTCATCCCAGGCCCTGGTACG 2223
Db   5459 GTTCGTCGGTGCTGATGCCGTGGTGTGTCAAGACCGGCGGGTCCTCGACATGTGGTACC 5518

Qy   2224 GCGGCAACGAGACGGGCGAACTCCATTGCCGACGTGCGGCTGTCTACGTCGGGTACA 2283
Db   5519 CGGGCGCAGGCGGGCGCCCCGAGGGCCGCGCGGCGCCCGCGTCAACCCGACGG 5578

Qy   2284 GCAAGCTGTCCCTCAGCTTCCCC-------AAGCGGCCTGCAGGACAAACCCGCGTTTC 2334
Db   5579 GCAAGCTCACGCACGCAGCTTCCCCGGCGCCGGAGTCGCGGGCGGACCCGA 5638

Qy   2335 TCAACTTCCCACGCCGGGGCGGGCCAGCCGTCGGACGTGCTACGTCGGGTACA 2394
Db   5639 CAAGCTACCCGGGGGTGCGACAACCAGAGCTACCGGAGGGGCCATCACGTTCGGGTACC 5698

Qy   2395 GGTACTACGGAGTTTGCCGACAAGGAGACCTCATTTCCCTTTGGCACGGCTGTCCTTACA 2454
Db   5699 GCTGGTTCGACAAAGGAGAACGTCAAGCCGCTCGTTCCCGTTCCGGTTCCCGGGTTACC 5758

Qy   2455 CCACTTTTGCCCTTTTCCAAATCTCTCTCACAAGGAC---GGCAAGCTGAGCGGTGT 2511
Db   5759 CCTCGTTCACGCAGCAGCGGCGCCCCGACCGTGCGTACGTTCCAGGGTGGTCTGAAGGTCA 5818

Qy   2512 CCCTCTCCGTGAAGAACACCGGCTTCCGTGCCCGGCCCACAGGTGGCCCAGCTCTACGTCA 2571
Db   5819 CGGTCACGGTCCGCAACGAGCGCGATGCGCGGGAGAGTCGTTCAGGCGGTACCTCG 5878

Qy   2572 AGCCCTCCAAGGCGCCCGTGAAGAATTAACCGACGGCTTCCGAAAAG 2631
Db   5879 GTGCCACCCGAACGTGACGCGCTCCGCAGCGGATCGAGAAGAGAGAGCTCGTGGGCTACACGGAAGG 5938

Qy   2632 TCGAACTGCAGCCCGGCGACGAAGGCGTGAC 2665
Db   5939 TCTCGCTCGCCGAGGACGGCGAAGACGGTGAC 5972
```

```
Qy  1384 GCTCTGCCGCACTCAGGGCGGCTACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGC 1443
         | |   ||       ||   ||     |    | |   |||      | ||| || |||
Db  5474 ACGGCGCAAGAGACCATCGGCCGGTCATCGGCCCCAGCGCGTCGAC-CCCAAGTCACCGGC 5532

Qy  1444 TCGAGACGCCGGCCCATCGTACACGCGTTCCTCCCATTCTTAGGCG 1503
         ||  || || |     |  ||    | ||   |   |  ||| |||
Db  5533 CTGGGCAGGCCCCACAGCGTCGTCCCGGACTCGCGGCACTCAAGGCC 5592

Qy  1504 AGCAGTGGCTCACGCCCCGACGGCGGCCATGCGCTGGAGGGTGTTCACGAGCCCC 1563
         ||  |  || |    |||   |||   | ||| |||     |  ||  ||| |||
Db  5593 CGCCGCGGGTTGCGGGTGCGACGGGTGACGTACGAGGCAG 5652

Qy  1564 CTGGTACCCCTACCGGCCGCAGCACATTGACGAGCTCTTCTTCACCAAGAGACGGACATCCACC 1623
         ||  || |  |||  || |||||||  || ||     |||  |
Db  5653 ATCCCGGCGGGGAACCTCAGC---------CCGGCGTTCAACC 5686

Qy  1624 TGGTGGACTACTACCACCCCCAAGCGGCGGCAGAGACACGTGGCTACGCCGACGACATGGAGGGCACGT 1683
         ||  || |  |  | |   |||  ||||| | || || |  | ||    ||  |    || |  ||  ||
Db  5687 AGGGCCCACAGCTCGAGCGCGGGCGAAGGCGCGCTTGTACGCGGCGCGCTTGACCGTGC 5746

Qy  1684 ACACCCGACGAGGACGACTGCACCTACGAGCTCGGGGGCGGGGGGCGGGGGGGCCGAGACACGTGGCTACGCCGACGACATGGAGGGCAAAGG 1743
         ||||| || | || ||||  |  |  | || | ||   | || ||   |   |  |  || |  ||
Db  5747 CCGCCAGGGCCGGGCTCAGGTTCTACGCCATCGGCGTCGTTAGGCGCCACACGGTGTGCAG- 5805

Qy  1744 CGTACGTAGACGACCAGCGTCGTCTGTCGACAAACGCCACCAAAGCAGAGTCCCCGGGGATGCCT 1803
         || |  |  |||   |   | ||   |   ||    |  |  ||||
Db  5806 -------------------------------------CTCGGGCAGCGCACA 5818

Qy  1804 TCTTCGGCTCCGCCACCCGCCGGAGGAGGACGCGGGGCCGCCGCGTCAAATCTCGTCAAGGGCAACACGT 1863
         |  | ||  ||   |  |||||| ||| | |||||     ||  |||   |   | | ||  ||| | ||
Db  5819 CCATCGAGGCCGGGTCAGGTTCTACGCCATCGGCGTGGTTACGCTGACCGAGG 5878

Qy  1864 ACAAGTTCAAGATCGAGTTCGGGCTCTCACCACCACTACACCCTCAAAGGGCGACCATCG 1923
         ||  | ||  |   | |  || || |   || |||||  |   |    |||  |  |
Db  5879 GCACCGACAAGCT----------------------CACGATCTCGGCGCTTCCGCGATGA 5914

Qy  1924 TCCCCGGCCACGGCTCCCGCTCCCGGTCGGGCGGCTGCGAAGGTCATTGACCAGGCCGAA 1983
         ||    || |||  |   |  |||     ||  ||    ||| | ||  ||| ||    |
Db  5915 GTGCCACCCGCTCTCCCTCCCGCTGGTGACGCGGGGTGAGCGGGCACCGACGA 5974

Qy  1984 TCGAAAAGTCCGTCGCCCTCGCGCCCGGTCGGAGGGCCCGGTACGGCGGCGAAGGCTGCTCCTTGC- 2043
         ||  ||  |||  |   | ||     |     | |  ||   |  ||    |  | |||  || |||
Db  5975 TCGCGAGGCGCGTCTGGAGTCGGCTCGGGCCGGAAGGCCCCGTACGGCGGTCGTCTTGC 6026

Qy  2044 ACGGCCGACTGGGAGAGACCGAGACCGGAGCGCATGAAAGCTCCCCGCGCGGTGCTGG 2103
         ||| || || ||  |    |  |   | |   ||  |||||  ||   |  || |||||
Db  6027 -CTACGACGACGGCACCGAGGGCGTCGAACCTTGTCGCGGGGTAGCGCAGG 6085

Qy  2104 ACCAGGTCATTGCCGACGTGGCGGCCCGCGCCGAACCAAAACACGTCGTTGCATTGCAGACGG 2163
         ||||  | | |||| | |||||| || |||   ||| |  |  ||| |||  | ||  | | ||
Db  6086 ACAAGCTGATCTCGGCGCTGCCGCGGCGGGAGCGCCAACCCGAACATCCGGTGGTTCCTCAACACCG 6145

Qy  2164 GCACCCCGGAGGAGAATGCCCTGGCTCGGAGCGGCGCGGACACCCGAACAGACCGAGCTGCTCGGTGACG 2223
         ||||    || || |  || ||  ||    || |  ||    | ||| | | |  |||| | |  ||  |
Db  6146 GTTCGTTCGGTGCTGCTGTTCCAAGACCCGCCGCGCGTCCCGTCGCTGTCTGGACCATGTGGTACC 6205

Qy  2224 GCGGCCAAGCGAGACGGGGCAAACTCCATTGGCGGAGTCGTGTTTGGGCGACGTACAAACCCCTCGG 2283
         ||| ||| ||| | ||||| || |  |||  ||  | | |  |  ||  ||  |  |  || ||||||
Db  6206 CGGGCCAACGGCAACTCCGGTGCGCGCGCGCCTGCTGCTTCTTACGCGCGGTTTC 6265

Qy  2284 GCAAGGCTGTCCCTCAGCTTCCC---------AAGGCGCTGCAGGACAAACCCCGCGGTTTC 2334
         ||  ||| |  || ||  | |||          ||||| |  |  |   |  || ||
Db  6266 GCAAGGCTCACGCACGAGCTTCCCGGCGAGGGCATCAGCGTCTACTCGGGGTACA 6325

Qy  2335 CAAGCTACCCGGCGTCGACAGAACGTACCAACGCCCTAACAGTATCTTG 2394
         ||  || ||||| | |||||   || | |  |||||    |   |  ||
Db  6326 GCAAGCTACGCGCGTCGCTTCCGCGGAGGCATCCAGGTTCGCGGTACC 6385

Qy  2395 GGTACTACGAGTTTGCCGACAAGGACGTCAAATTCCCCTTGGCGCACCGCGCTTGTCCTACA 2454
         |  ||   || | | |||||  ||| | |  |   ||||  |  || ||| |  || ||| ||
Db  6386 GCTGGTTCGGACAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACGACTGGTTGTACGGCA 6445

Qy  2455 CCACTTTGCCTTTCCAAATCTCTCCGTGTCTCCACAAGGAC---GGCAAGCTGAGCGTGT 2511
         || |  |||     ||  | |  | |  | ||  || |||     ||| ||| | |  ||
Db  6446 CCTCGTTCACGCAGCGGAGGCGCCCGACGGTCGTGGTCGTACGTCGCACGGGCTGGTGT 6505

Qy  2512 CCCTCCTGCCGTGAAGAACACCGGCGCCCCCGACGGTTCGCGTTGCTGCTGAGCGGCA 2571
         |||  |  | ||  | |  ||||| || || ||| |  ||  | | |  |  | | |||
Db  6506 CGGTCACGGTCGCGAACACGGGGCGGGCGGGCAGCGTCAAGGCTTCAGGGGCTACCTCG 6565

Qy  2572 AGCCCCTCCAAGCGCCACAGATTAACCGCCCCCGTCAAGGAGCTTCGCAAAGG 2631
         ||   | |   ||  ||  | | |  || |  ||| ||| | | | |   ||||
Db  6566 GTGCCAGCCCGACAGGTGACGGCTCCGACGCGAAGAAAGCTCGTGGGGCTACACGAGG 6625

Qy  2632 TCGAACTGCAGCCCGGCGGAGCACGAAGGCGGTGAC 2665
         |    ||| | || |||   |  ||| ||  || ||
Db  6626 TCTCGCGTCGCGCGGCGGGGGCGGAGGCGAAGACGGGTGAC 6659
```

```
Query Match       5.9%;   Score 175.2;  DB 10;   Length 13613;
Best Local Similarity   47.6%;
Matches 1016;  Conservative  0;  Mismatches 968;  Indels 150;  Gaps 11;

Qy   544 CGGCTCGGTTCCACATTCAACCAAAACTCTGCTCGGAAGAGGCAGGTAAGAATGAGATGGGGCAAAG 603
         |||  | ||  |  | |  |||  | | |||| |||| || || |||      |  |  | ||| |||
Db  4664 CCCTGGCCAGCAGCACCTTCGACGACGACCATGGCCGGCAGACTACGGCAAGGTCATGGGCGGGCGG 4723

Qy   604 AGGCCATGACGCTAAGAGTGCGCAATGTGATCCTCGGCCCCGACATTATCAACATGCAAAGGCTCCC 663
         | || |  |||    |||   || | || | ||  |||| |||  |  |  |||| |  |  ||| ||
Db  4724 ACGGTCGCGGCGCGCTCAACCAGGACACATGGTCCTGGGCCCGATGATGAACAACATCCGGGTGC 4783

Qy   664 CTCTCGGTGGACGTGGGCTTCGAGTCGATTGGGTGAGGAGATCGGGTTCCTGGCGGGCTTGGGAG 723
         |  | ||  ||   |||| | ||| || |  |  ||| | |  ||  | | || | | |  ||||
Db  4784 CGCACGGCGGCGAACTACGAGACCTTCAGCGAGGCTGCTACGGCTCTGCGGCGACACCG 4843

Qy   724 CTGGCGGCGCTCTCATCCGGGCCGCCATTCAGCACACTGGAGTGCAGGCTACGAGTCAACGCACTTTT 783
         | | ||| |  |||| ||  || || ||||| ||  ||||| ||| || |||| ||  || ||| |||
Db  4844 CGGTCGCCCAGATCAAAGGGCCATCAGGGGCATCAAGGGCTGGGCTGGCTGGACGAGAGCGC 4903

Qy   784 TGTGCAATGATCAGGAGGACAGAGCGGCCATCAATGGTGCAGAGCATCGTCACGGAGCGGGCTC 843
         | |  ||||  || ||||| |  || |||||  | |||  ||| |||  | || ||   | | |
Db  4904 CGGCTCGCCCAGATCAAAGGGCCATCAGGGGCATCAAGGGCTGGGCTGGCTGGACGAGAGCGC 4963

Qy   844 TCCGTGAAAAATCTACCGCCACTCCCGGTTCCAGATTGCTGTGCGGAGACGTCCCAGCGGGT 903
         | |||  || | |  ||| ||  | |  ||  || | |||| | || |||| || | |||
Db  4964 TCCGCCGAGATCGAGTTGCCCGGCGTTGCAAGCGCCAATGTCGACGAGCAGACGAGAGCG 5020

Qy   904 TCATGACGGCGGCTACAATGGCATCAATGGCCTGGTGTGCTGCGACGAGAACCCTAAATATCTTG 963
         ||  | || || |  |   ||| || |  |  ||  | | ||| |  | | || |  |  |  ||
Db  5021 TCATGTTGCCTACAACGGCCCTTCGCGCGGAAGCGCCGTCCTGCGGGGCAAACGACAGCTCCTCA 5080

Qy   964 ATGGGATGCGTTCGAAAGGAATGGGGTTGGGATGGCCCTAATCATGAGCGCGACTGGTTGTACGGCA 1023
         | ||| |||| | |||  |||  || |  |||||| ||||||||||||| |||| ||  | || | ||
Db     ... (continuation)
```

(Nucleotide sequence alignment — Query (Qy) vs Database (Db), positions shown at left and right of each line)

```
Qy  5081 ACAACGTCGCTGCCCACCGCAGTGGGCTTCCACGGCTTCGGGTGATGGTCGGACTGGCGTCGGCCA 5140
              ||||||||   ||||     |||   |||| |  ||| ||||  |||  |||   ||
Db  5081 ACAACGTGCTGCCCACCGCAGTGGGGTTTCTGGGAGTGTCGGACTGGCTCGGCCA 5140
Qy  1024 CATACAGTACCACAGAAGCCGTTGTGTGGCCAGCCTCGAGGATGCCGGACTTCCAC 1083
Db  5141 C---CCCGGGCACCGACGCCATCACCAAGGGCCTCGACCAGGAGATG 5184
Qy  1084 GCTTCCGAGGAGAGAAACACTCAAGTTCAAACGCCTTTATCCACGTCA 1143
Db  5185 ------GGCGTCGAGACTCCCGGAGACGTCCCGGAAGGGCGCCGGCCA 5236
Qy  1144 TTGACCAGAGGGCGCTAGGGAAGTTCTTCAGTTGTGCTGCCTCGGAGTGA 1203
Db  5237 AGTTCTTCGGGCGA---GGCGCTGAAGGAGGGCCACGGTTCCCGAGGCG 5293
Qy  1204 CGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCGAGAACGGCAGCTCTCCTCCGGA 1263
Db  5294 CCGTGACGCGGTCTGCGGGGAGCGGGATCGTCGGGCCACGATGGAAGGTCTGCTCCTCG 5353
Qy  1264 AGGTTGGCAACGAGAGGGCATCGTGCTCTGAAGACGAGAACAACGGTTCTGCCCTTGAGCA 1323
Db  5354 CCACTCCGGACGCGGCGGGCGGCGACAAGGGGTGTCCGCGGTTGTCCGCAAGG 5413
Qy  1324 AGAAGAAGAAGAAGCGTGATTGTTGTCGGCCCCCAACCGCCACATACCACGGCGGGAG 1383
Db  5414 TCGCCGAGAACGGCGGGTGCTCCTGCGCAAGCCAGGCGCCTGCGGCTGCCGTG 5473
Qy  1384 GCTCTCCCGCACTCAGGGCTTACTACGGACGCCCTTGACGGCGTCAGAGCAGC 1443
Db  5474 ACGCCGGCAAGAGCATCGGGGTCATCGGGCCCGACGGCGTCGAC-CCCAAGTCACGGC 5532
Qy  1444 TCGAGACGGCGGCCATCGTACACCGTGCGGGCCCTACACCACCGGTTCCTCCCATTCTTAGCG 1503
Db  5533 CTGGGCAGCGCGCCCACGTCGTCCCGGACTCGGCGGGCCACTCGACCATCAAGGCC 5592
Qy  1504 AGCAGTGCCTCACGCCGACGGCGCTCCGGGCCATGCGCCTGGGCGAGGGCGCCC 1563
Db  5593 CGCGCGGGGTGCGGGTGCGACGGTGACGTACGACGTCCGGGCACGTTCGAAGACCTTCGGACGGCC 5652
Qy  1564 CTGGTACCCCTAACCGCCAGCACATTTGACGAGCTCTTCTTCACCAACGACACC 1623
Db  5653 ATCCCGGCGGGGGAACCTCAGC----------CCGGCGTTCAACC 5686
Qy  1624 TGGTGGACTACTACCACCCCCAAGGCGCGAGACGTGTACGCGCGGACATGGAGGGGCACGT 1683
Db  5687 AGGGCCACCAGCTCGAGCCGGCGAAGGCGGGGCGCGTGTACCGACGGCGCCGGTGC 5746
Qy  1684 ACACGCGCCGACGAGGGACTGCACCTACGGACTCGCTTCGTCGCTGGCGAGGCGCAAAGG 1743
Db  5747 CCGCGGCACGGGGAGTACCGCCATCGCGGTCCCGTGCCACCGGTGCAG- 5805
Qy  1744 CGTACGTAGGACGACCAGCGTCGTCGTCGACAAGGCCACCAACGCAGGTCCCCGGCGATGCCT 1803
Db  5806 ------------------------------CTCGGCAGCGCCACA 5818
Qy  1804 TCTTCGGGTCTTCCGCCACCCGGCGGAGGAGGACGCGGCCGCATCACAATCTCTCACAGGGCAACACGT 1863
Db  5819 CCATCGAGGCCGGCGGGTCAGGTCTACGGCAAGGTGAGCAGCCCGCACCCCCGCGCTCGAAGCTGACCAAGG 5878
Qy  1864 ACAAGTTCAAGATCGACTTCGGCTCCACCACCTACACCCTCCAAGGGGGGACACCATCG 1923
Db  5879 GCACGCACAAGT-----CACGATCTTCGGGCTTCGCGGGACACCATCG 5914
Qy  1924 TCCCCGGCCGCCACGGCTCCTCCGTCGCGGGCTGCCAGGTCCCCGGCGATGCGCTTA 1983
Db  5915 GTGCCACCCCGCTCTCCCTCGAGCTCGGGCTGGTGACCGCCGGACGCCGACGA 5974
Qy  1984 TCGAAAAGTCCGTGCTGCCTCCACAAGGACGACGAGGTCATCATCTTCGCGGCGGCCTTA 2043
Db  5975 TCGCGAAGGCCGTGGAGTCGCGGCGGAAGCCCCGTTACGCGGGGGTCGCTCCTTCGC------ 6026
Qy  2044 ACGCCGGACTGGGAGACCGAGGGGCCGAGCAATGGAAGCTCCCGGGGGTGCTGG 2103
```

Query Match   4.7%;   Score 138.6;   DB 9;   Length 50937;

```
Db  6027 -CTACGACGACGGCCACCGAGGGGGTCGGTGGTCTCGAGCGTGTCGGACCTGTCGCTGTCGGGTACGCAGG 6085
Qy  2104 ACCAGCTCATTGCCGACGTGGCCGGGCGAGACCCAAACACCGGTTCGTCGTCATGCAGAGG 2163
Db  6086 ACAAGCTGATCTCTGGCTGTCGCGGACGCCACCAACCGGAACACGATCGTGGGTCCTCAACACCG 6145
Qy  2164 GCACCCCCGAGGAGAGGAGATGCCCTGGCCTCGACGCCACGACGCCGCCGGTCCAGCCTGGTACG 2223
Db  6146 GTTCGTCGTGCTGATGCCGTTGGTCGTCCAAGACACCCGGCGGTCCTGGGACATGTGGCTACC 6205
Qy  2224 GCGGCAACGAGACGGGCGACAACTCCATTGCCGACGTCGTTCTTTGCTTTGGCGGACTACACCCCTCG 2283
Db  6206 CGGGCCAGGCGGCGGGGCGGGCCGAGGGCCGCACCGGCTGCTCAGGTGTGCAACCTCATC 6265
Qy  2284 GCAAGCTCGTCCCTCAGCTTCCCC------------AACGCCTGCAGGACAACCCCGGCGTTTC 2334
Db  6266 GCAAGCTCACGCAGAGACGTTCCGGCCGGCGGGCGCCGACAGCGAAGCACGACGGTCGGGACCCGA 6325
Qy  2335 TCAACTTCCGCACCGAGGCCGGGGCGGGGCCGGGGGCGGTGTACGGCGGAGGAGCGTCTAGGTCGGGTACA 2394
Db  6326 CAAGCTACCCGGCGTCGACAAACCACCAGACGCCGGCTACCGCGACGTCCCACGTCGGGTACC 6385
Qy  2395 GGTACTACGAGTTTGCCGACACAAGGACGTCAATTTCCCCGTTTGGGCACGGCGCCTGTCCTACA 2454
Db  6386 GCTGGTTCGACAAGGAGACGTCAAGCCGGCTGTTCCCGGTTCGCACGGGCGGCGGTGTCGTACGTCA 6445
Qy  2455 CCACTTTGTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGAGCGTGT 2511
Db  6446 CCTCGTTCACGGCAGGACGCCCCGACCGTGTTCGGAGGGGTCTTACGGCGCAGGTCA 6505
Qy  2512 CCCTCTCCGTGAAGAACACACGGCGTCCGGTGCCGCGCGGCTACGTGGCCGCAGTCTCTACGTCA 2571
Db  6506 CGGTCACGGTCGCAACAGCGGGGAAGCGCGCGCCGCAGGAGGTCGTCGTACGTCGCTACCTCG 6565
Qy  2572 AGCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCCGTCAAGGCTTCGCAAGG 2631
Db  6566 GTGCCAGCCCGAAACGTGACGGCCTCCGCAGGCATCCACGTCGGGTACC 6625
Qy  2632 TCGAACTGCAGCCCCGGCGGCGAGACGAAGGCGGTGAC 2665
Db  6626 TCTCGCTCGCCGCGCGGGCGAGGCGAAGACGGTGAC 6659
```

Left column

Best Local Similarity 54.0%;   Pred. No. 1.2e-23;
Matches 354; Conservative 0; Mismatches 289; Indels 12; Gaps 3;

```
Qy  2048 CGACTGGGAGACCGAGGGCGCGGCAGGGCCGGCGAGCGCGGAGCATGAAGCTCCCGGGCGGTGCTGGACCA 2107
              ||||  |||||   ||  |||  ||||| | ||||  | |||    ||||  ||  ||||   |||||
Db  3417 CGACGACGGCTPCGGAGGACGGGGACGCGACGTGCGTGTCCCTGCCCGGCGCCAGGACGA 3358

Qy  2108 GCTCATTGCCGACGTGGCCGCGGCGAACCCAAACACCGTGCGTCGTTCATCGCAGACGGGCAC 2167
              ||  || |||||||||| || ||  | ||||  ||  |  ||| | |||||||||||
Db  3357 CCTGATCAGCGCGCGTGGCCGTGCGCGGGTCAACCCACGACCGGGGGCTC 3298

Qy  2168 CCCCGGAGGAGATGCCCTCGCGACCGCCACCCGCCCGGCCCCGTCAATCCAGGCTGGGTACGGGCGG 2227
              |||||||| ||| |||| ||||||  |||||||  ||  || || ||||||||||| |
Db  3297 CTCCCTCACCATCCCCTGGCTGAGGAAGACCCCGGTGCTCACCATGTGGTACCCGGG 3238

Qy  2228 CAACGAGACGGGGCAACTCCATTGCCGACGTGTCGTTGTTTGGGCGGCAA 2287
              ||||  ||||  | | || ||||||||||  ||| || |||||||||||
Db  3237 CCAGGCCGGGCCGGGACCACGCCCGGTCGGACCGGGACCCCAGGGAGGCCG 3178

Qy  2288 GCTGTCTCCCTCAGCTTCCC--------CAAGCGCCTGCAGGACAACCCGCGGTTTCTCAACTT 2341
              ||| || | ||  |||           ||||| ||||  ||||  | ||||
Db  3177 GCTGCAGCCAGACCTTCCCGGCGCCCAACGAGGGGCAGCGCCGGTTGCGCGACGGGCCCCAGGCG 3118

Qy  2342 CCGCGACCAGGCCCGGGGCGGCCACGGCTG---TACGGCGAGGACGTCTACGTTCGGGTACAGGTA 2398
              ||||||||||||  ||||||||||||         ||| |||||  ||  ||||| |||
Db  3117 CTACCAGTTTGCCGACAAGAGACGTCAATTTCCCCGTTGGCCAGGGCATCTACTCTACCACCAC 3058

Qy  2399 CTACCAGTTTGCCGACAAGAGACGTCAATTTCCCCGTTGGCCACGGCCTTCCTACCACCAC 2458
              || ||| ||| ||| || |||| | |||| |||| ||| ||  | |  ||  ||| |
Db  3057 GTACGAACAGGAGGACGTCCAGCCGTCCTCCGTTCCTCCGTCGACTGTGTCCT 2998

Qy  2459 TTTTGCCTTTTCCAATCTCTCCGTGTCTCACAA---GGACGGCGAAGCTGAGCGTGTCCCT 2515
              ||||| ||||  ||| || |||| ||  | |       ||| |||  ||   ||||||
Db  2997 CTTCGACTACCGGGGACCTGAAGGTGACGGCGCTCGACGTGCTTGCTT 2938

Qy  2516 CTCCGTGAAGAACACACGGCGTCCGTGCCCCAGCTCTACGGTCAAGCC 2575
              |||||  || ||| ||||||||| |||| ||  ||||||||||||||||
Db  2937 CACGCTGCGGCAACACCGGTACGCGCGGGAGTCCCGCGAGGTCTACGGTGGGGCC 2878

Qy  2576 CCTCCAAGCGCCCGCGAGATTAACCGCCCCGTCAAGGGCGTCAAGGGTTCGCAAAGGTCGA 2635
              |  ||    || ||||||| || || ||||  |||||  ||   |||||||||  || |
Db  2877 GTCCCGCACGTGCGGGTGGCGCAGCCCAAGCGGTCCCTGCCGGGCTGGAAGGTGA 2818

Qy  2636 ACTGCAGCCCGCCGAGACGAAGGCGGTGACAAATCGAGGACGACCAGGAGAAGTACGTC 2690
              |||||||| ||||||||||| ||||  ||||  ||||| | ||  |||||||||||
Db  2817 GCTGCGGCCGCGGGCCGGGCGGAGGAGCTGACCGTCGAGGCCGGAGGGCCCTC 2763
```

Right column

Query Match 4.6%;   Score 135.8;   DB 9;   Length 2166;
Best Local Similarity 53.5%;   Pred. No. 1.3e-23;
Matches 284; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

```
Qy  545 GCTCGGTTCCACATTCAACCAAAACTTCGCTCGAAGAGGCAGGTAAGATGATGGCAAAGA 604
             |||  || |||  ||  |||| ||| | ||| |  |||||||  |||||  ||| ||
Db  258 GCTCGCGTTCTPACCTGGAACAGAGACACCTTCTGGAGGAAGTGGGAAAAGCCATGGGGAGAAGA 317

Qy  605 GGCCATGCTAAGAGTCGCGCAATGTGATGTCCTGGCGCCGGGCCGACTATCAACAAGCCTCCCC 664
             ||||||| ||||| | | || ||||| ||| |||  || ||||||  || || ||||| |||
Db  318 GCTCGCGTTCTPACCTGGAACAGAGACACCTTCTGGAGGAAGTGGGAAAAGCCATGGGGAGAAGA 377

Qy  665 TCTCGGTTGGACGTGGTTCGAGTGGTTGAAGATCCGTTCCTGGCGGCGGCCTTGGGAGC 724
             | ||  ||||| |  ||  |||||  | ||| || | | ||  | || | ||||||||
Db  378 TCTTTGTTGGGAAGGAATTTCGACTACTCAGAGTGCTGTTCTCCGGGGTCGAAATGGC 437

Qy  725 TGCGGCTCTCATCCGCGGCCATTCAGGACCACTGGAGTGCAGCGCTACGGATCAAGCACTTTGTT 784
             |||| || |||  || | || ||||| | ||  || || |||||  || || |||||| |||
Db  438 TTCAGCCTTTGTCAAGGGAGTTCAAATCTCAAGGGTGGGCTGGACCTGCAATAAAACACTTTGT 497

Qy  785 GTGCAATGATCAGGAGGACGAGCAGCCGCATGAGTGGTACGGACATCGTTCCGAGGAGCGACCCT 844
             |||  ||||||||||||||  ||||||||  || || ||  || |||| | |||||| ||||||
Db  498 CGCGAACAACCAGGAAACGAGCAGGATGGTACTGACGCGATCGTTCCGAGCAGCGACCCT 557

Qy  845 CCGTGAAATCTACGCCACTCCCGTTCCAGATTGCGTCGGAGAGACTCCCAGCCGCGGTGCGTT 904
             ||| || |||| | || || ||||| ||| |||| | ||||  || ||||||||  || |||
Db  558 CAGAGAAATATATCTGAAGGTTTTTGAAATTGCTGTCAAAGAAGAAGCACCCCTGGACCGT 617

Qy  905 CATGACGGCGGTACAAATGGCATCAAATGGCGTGTCGTCGCAGCGGGAGCGAGAAACCCTCCCC 964
             |||||| || ||||||| |||| |||||  | ||||  ||| || || ||  ||  || ||||
Db  618 GATGAGCGCTTACAACAAACTGAATGGAAAATACTGTTCACAGGACGAATCGCTTTTTGAA 677

Qy  965 TGGGATGCTTCGAAAGGGAGGATGGGGTTGGGGATGGCGCCTAATCATGAGGACGACTGGGTACGGCAC 1024
             |||||| |||| ||| || || ||||  |||||||||||||| |||| |||||  ||||||||||||
Db  678 GAAGGTTCTGAGGAGGAATGGGGGATTTGGGGCCGGTTTGTTCGGTGATGAGCGGACTGGTACGGCGGGG 737

Qy  1025 ATACAGTACCACAGAAGCCGTTGTGGGCAGGCCTTCGACCTCGAGATGCCCGGG 1075
              ||  ||||| |||| || ||| ||  ||  ||||||||| ||||||| |||
Db   738 AGACAACCCTGTAGAACACGCTCAAGGCCGGAAACGATGATGATGCATGCCTGG 788
```

Lower left column

```
;  APPLICANT: Bylina, Edward
;  APPLICANT: Swanson, Ronald V.
;  APPLICANT: Mathur, Eric J.
;  APPLICANT: Lam, David E.
;  TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF
;  FILE REFERENCE: 09010-024006
;  CURRENT APPLICATION NUMBER: US/10/093,037
;  CURRENT FILING DATE: 2002-03-06
;  PRIOR APPLICATION NUMBER: US 09/910,579
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: US 09/134,078
;  PRIOR FILING DATE: 1998-08-13
;  PRIOR APPLICATION NUMBER: US 08/949,026
;  PRIOR FILING DATE: 1997-10-10
;  PRIOR APPLICATION NUMBER: US 60/056,916
;  PRIOR FILING DATE: 1996-12-06
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 5
;    LENGTH: 2166
;    TYPE: DNA
;    ORGANISM: Thermotoga maritima
US-10-093-037-5

Query Match          4.6%;  Score 135.8;  DB 9;  Length 2166;
Best Local Similarity  53.5%;
Matches  284;  Conservative  0;  Mismatches 247;  Indels  0;  Gaps  0;
                     Pred. No. 1.3e-23;

Qy  545 GCTCGGTTCCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGCAAAGA 604
            |||  | || | |||| || |  |  | || |  |   ||  || ||| | |  ||
Db  258 GCTCGCTTCTACCTGGAACAGAGACACTTCTGAAGAAGTGGGAAAAGCCATGGGAGAGA 317

Qy  605 GGGCCATCGCTAAGAGTGCCATGTAAGATCCTCGGCCCGACTATCACATGCAAACGCTCCC 664
            ||  |   || |  || |  |||  | |  || |||  | ||  | | | |||  |||
Db  318 AGTTAGGGAAATACGGTGTCGAATGTGCTTCTTGCACCTGCGATGAACATTCACAGAACCC 377

Qy  665 TCTCGGTGGACGTGGCTTCGAGTCGATGTTGTGAGGATCCGTTCCTGGCGGGCTTGGGAGC 724
           || ||  | ||  |  |  |  ||| | |  |  |   ||  | | | |||  ||
Db  378 TCTTGTGAAGGAAATTCGACTACTACTCAGAGAAATCCTGTTCCTTTCCGGTGAAATGGC 437

Qy  725 TGCGGCTCTCATCCGCGGCCATTCAGAGCACGTTGGGACGAGGTGGACTCAAGCACTTTTT 784
           |  || || ||| | | | ||  |   |   | || ||  || | |   | | |  |||
Db  438 TTCAGCCTTGTCAAGGAGGTTCAATCTCAAGGGGTGGGAGCCTCCATAAAACACTTTGT 497

Qy  785 GTGCAATGATCAGGAGGACAGGCCGCATGATCGGTCGTGCAGAGCGCGGGCGCTCT 844
           ||  | |  | | | |  ||  |  || || | |  || | | ||  || |  |
Db  498 CGCGAACAAACCAGGAAACGAACAGGATGGTAGTGGACACGATCGTTCGTCGAGCGCGCCT 557

Qy  845 CCGTGAAATCTACGACACTCCCGTTCCAGATTGCTGTGGCAGACTCCCAGCCGGGTGCGTT 904
           | ||||  | |  | || |||  | ||| || | |   ||| || | |||  || |||
Db  558 CAGAGAAATATATCTCAAAGGTTTTGAAATTGCTGTCTCAAGAAAAGCAAGACCCTGGACCGT 617

Qy  905 CATGACGACGGCGTACAATGGCATCAATGGCGTTGTCGTGCAGGGAGACACCTTAAAATATCTTGA 964
           || | | ||| | ||  | ||| | |  |  ||  | | ||| |   | | || | |  ||
Db  618 GATGACGCGCTTACAACAACTGAATGGAAATACTGTTCACAGAACGAATGGCTTTTGAA 677

Qy  965 TGGGATGCTTCGAAAGGAATAGGGGTTGGGATGGCCTAATCATGAGCGACTGGTTACGGCAC 1024
           ||| | || |||  | | | | ||| ||| | |  ||| | ||| | || | | ||
Db  678 GAAGGTTCTCAGGGAGGAGAATGGGGATTTGGCGGTTTCGTTGGATGCGGACTGTTTACGCGGGG 737

Qy  1025 ATACAGTACCACAGAAGACGCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGG 1075
            || |  |||| |  | | ||| || ||| | || ||| | |  | ||| |||
Db  738 AGACAACCCTGTAGAACAGCTCAAGGCCCGAAACGAATATGATCATGCCTGG 788

RESULT 10
US-09-790-399-7
;  Sequence 7, Application US/09790399
;  Patent No. US20020038000A1
;  GENERAL INFORMATION:
;  APPLICANT: Gold, Larry
;  APPLICANT: Tuerk, Craig
;  APPLICANT: Pribnow, David
```

```
;  APPLICANT: Smith, Jonathan D.
;  TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
;  FILE REFERENCE: NEX02/01-CON2
;  CURRENT APPLICATION NUMBER: US/09/790,399
;  CURRENT FILING DATE: 2001-02-22
;  PRIOR APPLICATION NUMBER: 09/197,649
;  PRIOR FILING DATE: 1998-11-23
;  PRIOR APPLICATION NUMBER: 07/829,461
;  PRIOR FILING DATE: 1992-01-31
;  PRIOR APPLICATION NUMBER: 07/739,055
;  PRIOR FILING DATE: 1991-08-01
;  PRIOR APPLICATION NUMBER: 07/561,968
;  PRIOR FILING DATE: 1990-08-02
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: Patentin Ver. 2.0
;  SEQ ID NO 7
;    LENGTH: 390
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;  FEATURE:
;    OTHER INFORMATION: Description of Artificial Sequence: Sequence
;    OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
;    OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-790-399-7

Query Match          2.8%;  Score 84.4;  DB 10;  Length 390;
Best Local Similarity 51.9%;
Matches  190;  Conservative  0;  Mismatches 176;  Indels  0;  Gaps  0;
                     Pred. No. 3.3e-11;

Qy  1661 GTACGCCGACATGGAGGGCACGTACACCGCCCACGAGGACTGCACCTACGAGCTCGGCCT 1720
            ||  | | || | | |||  |  | ||  || || ||  ||| |  | || |||  ||
Db     8 GGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 67

Qy  1721 CGTCGTCTGCGGCACGGCGCAAAGGCGTACGTAGACGACGACCAGTCTGCGTCGTCCGACAACGCCAC 1780
            | |  | ||  | |  |  |  |  ||| || |||||||||||||||  |||  |  ||
Db    68 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 127

Qy  1781 CAAGCAGGTCCCCGGCGATGCCTTCTTCGGCGTCCGCCACCGCGGAGGAGACGGGCGCCCAT 1840
             || || | |  | || |  | || ||  || | |||||||||||||||||||||||||
Db   128 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 187

Qy  1841 CAATCTCGTCAAGGGCAACACGTACACAGTTCAAGATCGAAGATCGGTTCGGCGTCCGCGTGCGCCACCACCTA 1900
             ||| | | |  |||  |  | ||  || || | |  |||||||||||||||||||||||||
Db   188 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 247

Qy  1901 CACCCTCAAGGGCGACACCATCGTTCCCCGGGCCACCGCCTCCCTCGCCGTGGCGGCTGCAA 1960
             | |||  || |  || ||| | ||  ||  ||||||||||||||||||||||||||||
Db   248 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 307

Qy  1961 GGTCATTGACGACGACCAGGCCGAAATCGAAAAAGTCCGTCGCCCTCCGCCAAGGAGCACGACGACGACCA 2020
             | | || ||||||||||||||||||||||||||| |   ||||||||||||||||||||||
Db   308 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 367

Qy  2021 GGTCAT 2026
            | |||
Db   368 CGCCAT 373
```

```
RESULT 11
US-09-748-033-2
;  Sequence 2, Application US/09748033
;  Patent No. US20020069431A1
;  GENERAL INFORMATION:
;  APPLICANT: Broadway, Roxanne M.
;  APPLICANT: Gongora, Carmenza E.
;  TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSIDASE AND THEIR
;  TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
;  FILE REFERENCE: 19603/3091
;  CURRENT APPLICATION NUMBER: US/09/748,033
;  CURRENT FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/172,003
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 8
```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1294
; TYPE: DNA
; ORGANISM: Streptomyces albidoflavus
US-09-748-033-2

Query Match   2.6%;   Score 78.2;   DB 10;   Length 1294;
Best Local Similarity   43.8%;   Pred. No. 2e-09;
Matches 396; Conservative   0; Mismatches 503;   Indels   6; Gaps   1;

```
Qy 1562 CCCTTGGTACCCCTAAACCGGCCACACATTGACGAGCTCTTTTCTTCACCAAGACGGACATGCA 1621
            ||| ||| | ||| |||||| |  | |  |  |||
Db  209 CCCCCCCTTCCGCACCCCGGCCGCGGCGCCTTCCACCCGGCCTTCACCCGCCAGGGCGCGGCT 268
Qy 1622 CCTGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTACGGCCGACGATGGAGGGCAC 1681
             |||
Db  269 GACCGCGGCCCCTCACCGCGGGCGCCGTTCCTCGGCGGCCTCGTGCGCCACGGCCCCGC 328
Qy 1682 GTACACCGACCGCCGACGAGGACTGCACCTACGAGACTCGGCCTTCGCGGCACGGGCAAA 1741
              || |  ||| |  |||| || | ||||  | ||||||   |
Db  329 GACCGGCGGGGCGGGAGGGGCCCCGGCGGGGCGCCCGTACCGGCCCCAGC 388
Qy 1742 GGGCGTACGTAGACGACCAGCTGTCGTCGTCGACAAGCGGCCACCAAAGAGGTCCCCGGCGATGC 1801
             ||| ||  ||||| || | || || | || |  ||| |||||||||   |||
Db  389 GGTGACCGGTTACTGGCGAGAACTTCAACAACGACCGTGCGACCCTCCCGCGACGT 448
Qy 1802 CTTCTTCGGCTCCGCCACCCGCGAGGAGACGAGGCCGCCATCAAATCTCGTCAAAGGGCCAACAC 1861
                ||||||| ||||   ||   |   |          | ||||  | | |||| |
Db  449 GCCGGACGGCCTACGACATCATCG--------CCGTCTCCTTCGCCGCAAGGGCAGCGGCGCCAAGC 502
Qy 1862 GTACAAGTTCAAGATCGAGTTCGGGTCGCCGACCCACACCCTTCAAAGGGCGACACCAT 1921
             ||| ||| || | |  ||| || |   ||| ||  || |  |
Db  503 GGGCGAGGATCACCTTCACCCTCGACTCGGTCGGGCGCCTACACCGACGAGCAGTT 562
Qy 1922 CGTCCCCCGGCCAGGGCTCCTTCCGCGGGGCGCGCCGTCATTGACCAGCCGA 1981
              ||| ||| || || | | || | |||  | || ||||||| ||||
Db  563 CCCGGCCGGGCAACTCGCCGCCCAAGACGGCCGAGCGGGGCGGGGTGGGCG 622
Qy 1982 AATGAAAAGTCCGTCGCCCTCGCCCAAGGAGCACGACCAGGTTCATCATCTGCGCGGGGCCT 2041
              | |||||||||  | | || |  || || |||| | |||||| |||| | || || |
Db  623 CGAGAAGGGGCGGTCCGGTCGCGTCAACGCACGCCCTCGGCCGCTTCGCCGACAGCAC 682
Qy 2042 TAACGCCGCGCACTGGGGAGGACCGGAGGGCCCACGGCCGCCGCCATGAAGCTCCCCGGCCGGTGCT 2101
               ||  ||  |  || ||  ||| ||  ||| |  | |  | | ||||  ||||| | |  |
Db  683 CTACGGCGGCTGATGGGAGGTACGGCTTCGAGCGGCGTCGGACGGGCGCTGAGAAACGGCCCT 742
Qy 2102 GGACCAGCTCATTGCCGACGTGGCCGCGCGGAACCAAACACCGTCGTCGTCGTCATGCAGAC 2161
              ||| || | ||  | |   | |||| | | | || |  ||||| |||| |
Db  743 CAACTCACCACTACATGACCGAGGCCCCTCACCAAGCTCCACGAGAGGGGGGACGGCCT 802
Qy 2162 GGGCACCCCGAGGAGGAATGCCCTGGCTGGGCCCGACGGCCCACGCCGTCATCCAAGGCCCTGGTA 2221
              ||| | | | || |  ||  |  |   || ||||   |  || |  |  ||| | |||| ||
Db  803 GGTGCTCACCATGGGCGGCCGCGACCATCGACATCGTCCCCGAGAACGAGGTGACTTCAA 862
Qy 2222 CGGGCGGCAACGAGACGGGGCAACTCCATTGCCGACGTGCTCGTTCTTTGGGACTACAACCCCTC 2281
              ||| ||||| | | | | || | | | | ||  ||  | | | || || | |||| ||
Db  863 GACGGCGGCTGGTCACGAAAGACTTCCTGACCGCCGTCAACATGCAGTACTACAACGCGG 922
Qy 2282 GGGCAAGCTGTTCCCTCAGCTTCCCCAAGCGCGTCGCGAGACAACCACCCGGGTTTCTCACCTT 2341
              |||| ||||| | | || || |  ||  || | || || ||||| |||| | ||
Db  923 CTCGATGCTCGGTCTGGACGCGCCAGTGTCTACGCGCAGGCGACGTCCAGGACTTCGGTGTCC 982
Qy 2342 CCGCACCGAGGCCCCAACGCTGGTGTCACTCTACGGCGTTGCTCACGACGTCTGGGGTACAGGTACTA 2401
              || || ||||| || | || | ||| | | || ||  |||| || | ||| | | |
Db  983 GCTCGCGCTTGACATCCAGCTGGAGAGAACGGTTCTCGAACGGCCTTCGAGACGTCTGGGTACTA 1042
Qy 2402 CGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCAGGCGGGTGGTACAGGTACTA 2461
              | || | ||  | |||  | |  ||| || | ||| |||| |||| | ||  ||
Db 1043 CGCCTCCCGAAGGCGCGCGGCCTAGTCGAGGCCTCGGTGGTCAACGACGCGCGCT 1102
Qy 2462 TGCCT 2466
             | ||
Db      GGACT
```

Db 1103 GGACT 1107

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Streptomyces albidoflavus
US-09-748-033-6

RESULT 12
US-09-748-033-6
Sequence 6, Application US/09748033
Patent No. US2002006943A1
GENERAL INFORMATION:
    APPLICANT: Broadway, Roxanne M.
    APPLICANT: Gongora, Carmenza E.
    TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSIDASE AND THEIR
    TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
    FILE REFERENCE: 19603/3091
    CURRENT APPLICATION NUMBER: US/09/748,033
    CURRENT FILING DATE: 2000-12-22
    PRIOR APPLICATION NUMBER: 60/172,003
    PRIOR FILING DATE: 1999-12-23
    NUMBER OF SEQ ID NOS: 8

Query Match   2.6%;   Score 76.2;   DB 10;   Length 1107;
Best Local Similarity   45.7%;   Pred. No. 5.9e-09;
Matches 390; Conservative   0; Mismatches 448;   Indels 15; Gaps   3;

```
Qy 1467 GTCGGGCGCCTACACACCACCGTTCCTCCCATTTCTTAGGGCGGAGCAGTGCCTCACGCCCGACGGC 1526
              ||| | ||  || | ||| |   |  || || | | ||| | || || | | |||| ||||
Db  118 GGGCGGCAAGTGCACCATCGGGTGACAGCTTCGGCCGCCTTACCGGCGCCTAACCAG 177
Qy 1527 GCTCCGGGGCCATGCGCTGGAGGGGTCTTCAAACGAGAGCCCCCGTTGGTACCCCTAACCCCAGCAC 1586
             |   |  ||  ||||  | || |  ||||| || | |  | | |  ||||||  ||
Db  178 GAGTCGGTCGACGGCGTGCCCGACACCTTGGACAGCCGCCTCGGCGCCAACTTCAACCAG 237
Qy 1587 AATTGACGAGCTGTCTTCTTCACCAAGACGACGACATGCACCTGGTGGACTACTACCACCCAAG 1646
             | ||  | |||| | ||  | |||  ||||  ||||| || |  ||||  ||| | ||
Db  238 CTCCGCAAGCTCAAGAGCCCAAGTACGCCGCCACATCAAAGGTCCCTCTGGGTCCTTCGGGGCTGG 297
Qy 1647 GCGGCGAGACACGTGTACGCCGACACATGGAGGGGCGACGTACACCGCCGACGAGGACTGCACC 1706
             || |   || |  || |  || | |||| || | | | |  || || || ||  |  |
Db  298 ACCTGGTCCGGCGGCGGCTTCACCGACGCGTGAGAAGAACCCGGCCCCTCGCCAAGTCCTGC 357
Qy 1707 TACGAGCTCGGCCTGCTCGTCGTCTGCGGCACGGCCAAAGGGCTACGTAGACGAGTCGTTC 1766
              | || | | |   | | | | | | ||| |  || ||||  | | |   ||  ||| ||
Db  358 CACGACCTGG---TCGAGGACCCGCGCGTGGGCCGACGTCGTTCGACGGCATCGACCTGCGAC 414
Qy 1767 GTCGACAACCGCCACCAAAGCAGAGGTCCCCGGCGATGCCTTCTTCGGCTCCGGCGGGGGAG 1826
             || || | |   ||||  || || | || |||| | ||| | || || || | | ||
Db  415 TGGGAGTACCCGACAACCGCCGCCCTCGGCCGTCAGCTGCGCGACAGCTCCGGTCCGGGGCTGAAG 474
Qy 1827 GAGACGGGCGCCGATCAAATCTCGTCAAAGGGCAACACGTACAAAGTTCAAGATCGAJTTCGGC 1886
             | | || || ||| || || | |||||| || |  || | ||| || |  ||| |  ||
Db  475 AACATGGTCACCAGGCGAGATGCCGCGCCCAGTTCGGCACGTCGGTTCGACGGCATCGACCTCACC 534
Qy 1887 TCCGCACCCACCTACACCCTCAAAGGAGGCGACAACCGTCGTCCCCCGGCGCTCCCTCGGC 1946
             || ||| || | || || | |   | | |  | | || | || ||| ||| | |||
Db  535 GCCGACGCCAGCGTCCGGCCGCCAAGGCTCGACCGGCCGCCGCCGCCGCCCAGTAC 594
Qy 1947 GTCGGCGGCTGCAAGGTCAT-------TGACGACCAGGCCGACGGCGAAAATCGAAAAAGTCCGTCGCC 2000
              | ||| |  ||   |  ||       | || | || |  | ||| || || ||| | ||||
Db  595 TTCGACTGGGTTACAACGTTGATGACGTACGACTTCTTCGCGGCGCCCCTGGGACAAGAACCGGCCG 654
Qy 2001 CTCGCCAAGGAGGACGACGACCAGGTCATCATCATCTGCGGGGGCCTTACGCCGACTGGGGAGAGACC 2060
              |  | | ||  || | ||| |   | || || | ||  |  |  || | || || |
Db  655 ACCGGCGCCCCACTCGGCCCTGAACTCCTACACGGCCTTGGGGGCATCCCCAAGCGCGACGACTCCCACTCG 714
Qy 2061 GAGGGCGCCCGAGCCGGAGCAATGAAGCTCCCCGGCGGGGGGGGGGTGCTGGACACCAGTAC 2120
              |  ||  | || || ||| | | | || || | | | |  | | ||  | | |
Db  715 GCCGCGCCCCCATCGCCAAGCTCAAGGCCGGAGGGGAGCGAAGGGCGTCCCTGCTCGGCGGC 774
```

RESULT 13

```
US-09-748-033-4
Sequence 4, Application US/0974803
Patent No. US20020694431A1
GENERAL INFORMATION:
  APPLICANT: Broadway, Roxanne M.
  APPLICANT: Gongora, Carmenza E.
  TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSIDASE AND THEIR
  TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
  FILE REFERENCE: 1903/3091
  CURRENT APPLICATION NUMBER: US/09/748,033
  CURRENT FILING DATE: 2000-12-22
  PRIOR APPLICATION NUMBER: 60/172,003
  PRIOR FILING DATE: 1999-12-23
  NUMBER OF SEQ ID NOS: 8
  SOFTWARE: PatentIn Ver. 2.1
  SEQ ID NO 4
  LENGTH: 2712
  TYPE: DNA
  ORGANISM: Streptomyces albidoflavus

Query Match    2.68%;
Best Local Similarity   45.7%;
Matches  390;  Conservative   0;  Mismatches  448;  Indels   15;  Gaps   3;

Score 76.2;  DB 10;  Length 2712;  Pred. No. 9.1e-09;
```

```
Qy   1467 GTCGGGCGGCCTACACCACCGTTCCTCCCATTCTTAGCGACGCAGTGCCTCACGCCCGACCGC 1526
Db   1685 GCCGGCAAGTCCACCATCGGTGACAGCTTCGCCCGCTTACGACAAGGCGTACACCGCCGCC 1744

Qy   1527 GCTCCGGCATGCGTGGAGGGTCTTCAACGAGCGCCCCTCGGTACCCGTAACCGCCAGCAC 1586
Db   1745 GAGTCGGTCGACGGCGTCGGCCGACCACTGGCGCCGCCGCAACTTCAACCAG 1804

Qy   1587 ATTGACGAGCTGTCTTCTTCACCAAGACGGACATGCACCTGGTGGACTACTACCACCCAAG 1646
Db   1805 CTCCGCAAGCTCAAGGCCAAGTACCCGCACATCAAGGTCCTCTTGGTCCTTCGGCGGCTGG 1864

Qy   1647 GCCGGCAGACACGTGGTACCGCGACATGGAGGGCGACGTACACCGCCGAGGAGGACTGCACC 1706
Db   1865 ACCTGGTCCGGCGCGGCTTCACCGACGCCCTGAGCAACCGGCGCGCTTTCGCCAGAGTCCGTGC 1924

Qy   1707 TACGAGCTCGGCCTCGTCGTTCGCGGACGCAAAGGCGTACGTAGACGACAGCGTCGTC 1766
Db   1925 CACGACCTGG---TCGAGGACCCGCGCTGGGGCGCGATGCCCCATGTCGACCTTCGAC 1981

Qy   1767 GTCGACAACGCCCACCAAGCAGGTCCCCGGCGATGCGTCGCTTTCTTCGCCGTCTCCGCCGCGAG 1826
Db   1982 TGGGAGTACCGGACCGAAGCCGTTCGAGTCTCCGGTCCGGCGCGCGCGGCTGAAG 2041

Qy   1827 GAGACGGGCGCGCATCAAATCTCGTTCGCAGGGCCACAACTGTTCAAGAATCGAGTTCGGC 1886
Db   2042 AACATGGTTCCAGGCGGATGCGCGCCCACGTTCGGCACCGACTTCGGTCACCGGCCGCGCC 2101

Qy   1887 TTCCGCACCCACTACACCCTCAAGGGGCACACCATCGTTCCCGGCCACCGGCTCCTCCCGC 1946
Db   2042 ...
```

RESULT 14

```
US-09-815-242-7960
Sequence 7960, Application US/09815242
Patent No. US20020615569A1
GENERAL INFORMATION:
  APPLICANT: Haselbeck, Robert
  APPLICANT: Ohlsen, Karl L.
  APPLICANT: Zyskind, Judith W.
  APPLICANT: Wall, Daniel
  APPLICANT: Trawick, John D.
  APPLICANT: Carr, John
  APPLICANT: Carr, Grant J.
  APPLICANT: Yanamoto, Robert T.
  APPLICANT: Xu, H. Howard
  TITLE OF INVENTION: Identification of Essential Genes In
  TITLE OF INVENTION: Prokaryotes
  FILE REFERENCE: ELITRA.011A
  CURRENT APPLICATION NUMBER: US/09/815,242
  CURRENT FILING DATE: 2001-03-21
  PRIOR APPLICATION NUMBER: 60/191,078
  PRIOR FILING DATE: 2000-03-21
  PRIOR APPLICATION NUMBER: 60/242,578
  PRIOR FILING DATE: 2000-05-26
  PRIOR APPLICATION NUMBER: 60/206,848
  PRIOR FILING DATE: 2000-05-23
  PRIOR APPLICATION NUMBER: 60/207,727
  PRIOR FILING DATE: 2000-05-23
  PRIOR APPLICATION NUMBER: 60/253,625
  PRIOR FILING DATE: 2000-11-27
  PRIOR APPLICATION NUMBER: 60/257,931
  PRIOR FILING DATE: 2000-12-22
  PRIOR APPLICATION NUMBER: 60/269,308
  PRIOR FILING DATE: 2001-02-16
  NUMBER OF SEQ ID NOS: 14110
  SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 7960
  LENGTH: 1914
  TYPE: DNA
  ORGANISM: Pseudomonas aeruginosa
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (1)...(1914)
  US-09-815-242-7960
```

```
Db   2102 GCGGACAGCCAGCTCCGGCGGCGGCAGCCGCAAGCTGGAGGGGCGGGGACTGCTACGAGCAGGC 2161
Qy   1947 GTCGGCGGCTGCAAGGTCAT------TGACGACCAGGCCGAGAAATCGAAAAGTCCGTCGCC 2000
Db   2162 TTCGACTGGTACAACGTTGATGACGTTACGAGACTTCTTCGGCCGCTGGGACAAGACGGCCCG 2221

Qy   2001 CTCGCCAAGGAGGACCACGACCAGGTCATCACTTCGGCGGGGCCTTAACGGCGACCGTGGGAGACC 2060
Db   2222 ACGGCCGCCCACTCGGCGCCTGAACCTCCTACAGGCCGCATCCCCAAGGCCGAGTTCCACTCG 2281

Qy   2061 GAGGGCGGCCGAGGACGCCGGCGAGCCATGAAGCTCCCGGGCGTGGTGCTGGACCAGGTCATTGCCGAC 2120
Db   2282 GCCGGCGCCATCGCAAGCTCAAGGCCGAGAGGGCGTCCGGGCGGAGCCAGTCCTTGCCTCGGC 2341

Qy   2121 GTGGCCGCGGGCGGGAACCCAAACACCGTCGTCGTCATGCAGA------CGGGCACCCCGAG 2174
Db   2342 GTCGGCGGCTGCAAGGTCAT------TGACGAC... 2401

Qy   2175 GAGATGCCCTGGCCTCGACGCGCACCGCCTGGACCGGCGTTCGTGGTCTGTTGTCAGTGCAGA 2234
Db   2402 ACCGGCGCCCGCGCCACCGGCCACCTACGGAGCCATGAGGGTCCTCAGGAAC 2461

Qy   2235 ACGGGCGAACTCCATTGCCGACGTCGTCTTTGGGGACTACAACCCCTCGGGCGAAGCTGTCC 2294
Db   2462 ACCTGCGCCCCGCCACCGGACCACGGTCGGGCGCCACGCCGTACGGCCGCGCAAGTGGGCGCAGCAACTGG 2521

Qy   2295 CTCAGCTTCCCCA 2307
Db   2522 TGGAGCTACGACA 2534
```

```
Query Match      42.6%;  Score 73.6;  DB 10;  Length 1914;
Best Local Similarity   42.6%;  Pred. No. 3.4e-08;
Matches  452;  Conservative   0;  Mismatches 599;  Indels   9;  Gaps   1;

Qy  1573 CTAACCGCCAGCACAITTGACGAGCTCTTCTTCACCAAGACGGACAATGCACCTGGTGGACT 1632
              ||||| |  ||||||     ||  |   | |  | | |      ||||| ||    ||||
Db   161 CCAAGCGCCAGCGGTCACCAACCCGCAGAACACCCTGTATGCGGGTGAAGGCGCCTGATCG 220

Qy  1633 ACTACCACCCCAAGGCGCACGTGGTACGCCCACATGGAGGCGCACGTACACCGCCG 1692
             |||||   |||||||||  | | ||| |  | || | | ||  ||  |||
Db   221 GCGCTGCCTTCGAAGAGAACGTGGTTCAGAAAGACATTCAGATGGTTCCGGTACAGCATCG 280

Qy  1693 ACGAGGACTGCACCTACCAGTCCTCGGCCTCGTCGTCGTCCGCACGGCGTACGTAG 1752
              |||  || ||    |||||  ||| |   |||  ||| | ||| || |||||| |||
Db   281 TCAAGGCGCAACACGGCCTGGGTGGAAGTGAAGGGCCAGAAAGATGCGCGCTCCGC 340

Qy  1753 ACGACCAAGCTGTCGTCGTCGACAACGGCCACCAACGAGGTCCCCGGCGATGCCTTCTTCGGCT 1812
              |||| ||| |  || |||||| |||  |||  |  |||  ||  | ||| |  || |  |||
Db   341 AGATTTCCGCCGAAGTGCTGAAGAAAATGAGAAGACCGGCAACACGTACAAGTTCA 400

Qy  1813 CCGCCACCCGGAGGAGACGAGGCGGCCGCATCAATCTGTCAAGGGCAACAAGTTCA 1872
             ||| ||| |||| | |  | || | ||  ||  |  || || | ||  ||  ||
Db   401 CGGTCACCGAAGCGGTGATCACCGTTCCGGATCATCAACACACCATCGTTCCCGGCC 460

Qy  1873 AGATCGAGTTCGGCTCCCGACCTACACCCTCAAGGGCGACCGACAGCTCGTGTCATGCGACGG 1932
              |||||  | |||  ||| |  | ||  || || | || ||  ||  | |||| |  |||||
Db   461 CCAAGGACGCGGGCGCCATCGACGTCAAGCGGATCATCAACGAGCCGACCG 520

Qy  1933 ACGGCTCCCTCCGGCTGCAAGGTCAITGACGACCAGCCGGAAATCGAAAGT 1992
              ||||| ||  ||  |  |||| |||||||| ||||| || || |||||  |
Db   521 CGGGCGGCGTGGCCTTACGCGGCCTGAACAGGCCACACCCGTTCAAGGTGG 580

Qy  1993 CCGTCGCCCTCGCCAAAGAGCACAGACCAGGTCAITCATCTCACGGCCGACT 2052
              ||||| | |  ||||  |  | ||  | |||||  ||  |  || |  ||||
Db   581 AGCTGGGCGCGGTTACCTTGCAGTGTCCGTGATCGGAGATCGGGCCGAGC 640

Qy  2053 GGGAGACCGAGGGGCGCCGCGACCGCCGGAGCATGAGAGCTCCCCGGCGGGTGCTGGA 2104
              ||  |||  || | ||| | |  |  |||||   ||||| |||||| | | ||||
Db   641 ACCAGTTCGAAGTGCTGGCCACCAACGGCGACACCTTCCTGCGGCCGGACAGACTTCGACA 700

Qy  2105 -CCAGGTCAITGCCGACGTGGCGCGCGGACCAACCCAAACACCGTGTCGTGTCATGCGACGG 2163
              ||||| | | | || | | | | ||  |||  | |  |||  |||  |  | |||||
Db   701 TCCGGCTTGATCGACTTACCTCGTCGTCGACGAGTTCAAGAAGGAAAGCGGCATCAACCTGAAGG 760

Qy  2164 GCACCCCCCGGAGGAGAITGCCCTTGGGCGTCGAGCGCCACGCCCCCGCGGCTGGTACG 2223
              |||  |  | ||||||  |  |  |||   ||| |  ||  | |  | |   | ||||
Db   761 GCGACCCGGCTGCCATGCAGGTCTCAAGGAGAAGCGCCCGGTGGTGGCGTGT 820

Qy  2224 GGGGCAACGAGAGACGGGCAACTCCATTGCGCGACGTCGTCTTTTGGGCACTACAACCCCTCG 2283
              ||| | ||  |  ||  |||  | ||  || | ||  |||  | | |||  |||  |  ||
Db   821 CCTCGACCCAGCAGACGGCTACGTGACGTCGGAGCGGGTCCGA 880

Qy  2284 GCAAGCTGTCCCTCAGCTTCCCCACAAGCGCTGCAGCACAACCCCGCGTTTCTCAACTTCC 2343
              ||  | ||  |||  ||   | ||  | ||  |  || |||  |   | |  | ||||| |
Db   881 AGCACCTGAACGTACAAGGTTTCCGGCGCCAAGCGGCCAAGCTGAGTGTCGAGAACTGGTGC 940

Qy  2344 GCACCGAGGCGCGGGGCGACGCTGTACGGCAGGACGTCTACAGGGGTGTTACGGCTACTACG 2403
              || |||||| |  | |   |  ||| |||  |   ||| |   ||| | | |||  |  ||
Db   941 AGCCGCACATCGACCATCGAGCCGTGGCTGGGGTCGGTGCTGGATGCGTGCGCCTGGATGTTTCCGGATA 1000

Qy  2404 AGTTTGCCGACAAGGAGCGTCAAATTTCCCCTTTGGCCACACGGCGTGTCCCTCTCACCACTTTTG 2463
              || | |||  ||  |||||| |  || |||  || ||||  || |  |  || || |||  ||||
Db  1001 TCCAITGAAGTGATCCGTGCTGGTGCGTGTCCCTCTACAGACCGTTG 1060

Qy  2464 CCTTTTCCAAATCTCTCGTGTTCACACAAGGAGGGCGGGAAGAGACCGCAAGCTCTCTCCGTGA 2523
              |||   | |||| | | | |  |  |||||| || || ||  |  |||   || || |||||
Db  1061 CCGAGTTCTTTGGGCGAAGGCGCCGCAAGGACCTCACCGGCGTGTCTG 1120

Qy  2524 AGAACACCGGCTCCGTGCCCGGGCCACGTGGCGGGCCGACGTTGCTCTGCTCGACG 2583
              ||| | ||||| || | |||| | |||  | || |  |||  || |||| ||||||
Db  1121 GTGCGCGCATCCAGGGCGCGGGTCCTGTCGGGCCGACGTTGCTCTGCTCGACG 1180

Qy  2584 CGGCCAAGAGATTAAACCGCCCCGTTCAAGGGGCTTCAAGGGGCTT 2623
              ||| ||  || | | ||   | || |  ||  |||| | ||||||| |
Db  1181 TGACCCCCCGGCTGACCCTCGGCCATCGAGAGACCCTGGGCGCGGCGT 1220
```

```
Query Match      43.0%;  Score 73.2;  DB 10;  Length 2541;
Best Local Similarity   43.0%;  Pred. No. 4.9e-08;
Matches  464;  Conservative   0;  Mismatches 608;  Indels   6;  Gaps   2;

Qy  1593 GAGCTCTTCTTCACCAAGACGGACAATGCACCTGGTGGACTACTACCACCCCAAGGCGCA 1652
              ||| | |  ||    |  || ||  |  || ||  |  |  | || ||||  || |
Db   439 GACCGCGGCGAGATCAAGAACACTGCAGCTTCGAGGTGACCACAGCCAACAAGATG 498

Qy  1653 GACACTGGTACGCCCACATGGAGGCGCACGTACACCGCCGACGAGGACTGCACCTACGAG 1712
              ||| | || ||| | || |  ||  ||| |  |||| | |||   |  |  | |||
Db   499 CAGAAGGAGTACGCCCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACGACAAACC 558

Qy  1713 CTCGGCCTCGTCGTCGTCCGCACGGCGAAAGCGTACGTAGACGACCGACGTCGTCGAC 1772
              ||| | |  ||  |   |  ||    | | | ||| |  ||| |  ||  | | ||
Db   559 AGCTACAAGCTGATCAACTGACAACCAGCTGATCACCCAGGCCTGCCCAAGGTGAGC 618

Qy  1773 AACGCCACCAAGCAGAGTCCCCGGGCGATGCCTTCTTCGGGCTCCGGCGAGAGACG 1832
              ||| || ||  |  | || |  | | |  | ||  |  |  ||| | |  || ||||
Db   619 TTCGAGCCCATCCCACTACTGCCCGCGCGCCATCTTCGCCATCCTGAAGTGCAAC 678

Qy  1833 GGCGCGCATCAAATCCGTCAAGGGCAACGACTACAGTTCAAGATCGAGTTCGCGGTCCGCA 1892
              ||| | |||   | || |||| | ||| |  |  ||| |  ||  | |||  | ||| ||
Db   679 GACAAGAAGTTCAAACGGCAGCCGGTCACCAAACCTGTGAGCGTGCAGTTGCACCCAC 738

Qy  1893 CCCAACCTACAACCGCTCAAGGCGGCGACACCAATCGTCCCCGGGCGCCTCGCGTGGCC 1952
              || | |  | || | | ||||| | |  |  |||| |  ||||| || | | |  |||||
Db   739 GGCATCCGGCGGCTTTTGCGTGGGTGAGCACGTGCTCTGAACGGCGACCGCCTCGAGGAGGC 798

Qy  1953 GGCTGCCAAGGTCATTGACGACCAGGAGGCGGAAATGAAAAGTCCGTCGCCCTCGCCAAGGAG 2012
              ||| | | |||  | ||| ||| |||  ||| ||  | | || ||  | | | | |  | |||
Db   799 GTGGTGATCCGCAGCGAGAACTTCACCGACGACCAAGACCATCGTGCCAGCTGAAG 858

Qy  2013 CACGACCCAGGTCATCATCTGCGCGCGGCCTTAACGCGGACTGGAGAGGCGGCCGAC 2072
              ||| || |||   | | |||| |   | | | ||   |  | || ||| | || ||
Db   859 GAGAGCGGTGAGATCAAATGCACCGCCCAACAACACCAAGAGCATCACCATC 918

Qy  2073 CGCGCGGAGCAATGAAGCTCCCCGCCGTGCTGGGACCAGCTCAITGCCGTGGCGCGCG 2132
              || | |||| ||| |||| |  ||   ||  |||  ||  | |  ||| |  ||| |||
Db   919 GGCCCCGGCGCGGCCCTTCTACGCC---ACCGGCGACATCATCGGCGAGGAGATCGGCGTGCG 975

Qy  2133 AACCCAAAACACCGTGTTCGTCGTCATGCCAGACGACGGCGACATCACCATTGCC 2192
              || | | || ||  | ||  |  || | |||  |  ||| || | |  |  ||||
Db   976 CACTGCAACATCACGCGGCGAGAAGTGGACAACAACCCCTGAAAGGCAAG 1032

Qy  2193 GCCACGCCCCGTCAITCAGGACCTGGTACGGCCGCAAGCTCCAITGCC 2252
              ||  | | |  | |  | || |  | || | ||  | |  | || ||
Db      
```

```
Db  1033 CTGCAGGCCCAGTTCGGCAACAAGAGACCCATCGTGTTCAAGCAGCAGCGGCGGGGGGACCCC 1092
Qy  2253 GACGTCGTCTTTTGGCGACTACAACCCTCGGGCAAGCTGTCCCTCGAGCTTCCCAAGCGC 2312
          ||  ||||
Db  1093 GAGATCGTTGATGCACACCTTCAACTGCGCGGGCGCGAGTTCTTCTACTGCAACACGACCCAG 1152
Qy  2313 CTGCAGGACAACACCCCGCGTTTCTCAACTTCCGCACCGAGGCGCGCACGCTGACGGC 2372
          |||  ||||
Db  1153 CTGTTCAAACAGCACCTGGGACAAACACCATCGGCCCCAACAACACCAACGGCACCCATCACC 1212
Qy  2373 GAGGACGTCTACGTCGGGTACAGGTACACCGAGTTTGCCGACAAGGACGTCAANTTTCCCC 2432
          |||  || ||||   |  |
Db  1213 CTGCCCTTGCCGCCATCAAGCAGATCATCAACCGGCTGGGGGGGCAAGGCCCATGTACGGCCCCCC 1272
Qy  2433 TTTTGGCCACGGCCCTGTCCTACACCACTTTTGGCCTTTTCAAATCTCTCCGTGTCTCACAAG 2492
          ||  |||||| || +
Db  1273 CCCATCCGGCCGGCCAGATCCGCTGCAGCAGCAACATCACCGGCCTTGCTGCTGACCCGGCGAC 1332
Qy  2493 GACGGCAAGCTTGAGCGGTGTCCCTCTTCCGTGAAGAACACCGGGCTCCGTGCCCGGCGCACAG 2552
          |||||||||  ||  |
Db  1333 GGCGGCAAAGGAGGATCAGCAACACACCGAGAATCTTCCCCCCGCGGGGCGACATGCGGC 1392
Qy  2553 GTGGGCCCAGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGAATTAACCGCCCGTCAAGGAG 2612
          ||   ||
Db  1393 GACAACTGGCGCGCAGCGAGCTGTACAAGTACAAGGTGGTTGAAGATCGAGCCCCTGGGCGTG 1452
Qy  2613 CTCAAGGGCGTTCGAAAAGTCGAAACTGCAGCCCGGCGAGACAGAAGGCGGGTGACAATCG 2670
          ||   ||||  ||  |||||| + |||||||   ||||||
Db  1453 GCCCCCACCAAAGGCCCAAGCGCCGCGGGTGGTGCACGCGCGAGAAGCGGCCGTGACCCTGG 1510
```